



This work is protected by copyright and other intellectual property rights and duplication or sale of all or part is not permitted, except that material may be duplicated by you for research, private study, criticism/review or educational purposes. Electronic or print copies are for your own personal, non-commercial use and shall not be passed to any other individual. No quotation may be published without proper acknowledgement. For any other use, or to quote extensively from the work, permission must be obtained from the copyright holder/s.

COMPUTATION APPLICATIONS IN ARCHAEOLOGY

A Thesis submitted in the University of Keele
In fulfilment of the requirements for the
Degree of Doctor of Philosophy (Ph.D.)

By

John Denis Wilcock

M.A.(Oxon.), B.Sc., C.Eng., M.I.E.E., M.B.C.S.

September 1972



IMAGING SERVICES NORTH

Boston Spa, Wetherby
West Yorkshire, LS23 7BQ
www.bl.uk

BEST COPY AVAILABLE.

VARIABLE PRINT QUALITY



IMAGING SERVICES NORTH

Boston Spa, Wetherby
West Yorkshire, LS23 7BQ
www.bl.uk

**TEXT BOUND CLOSE TO
THE SPINE IN THE
ORIGINAL THESIS**



IMAGING SERVICES NORTH

Boston Spa, Wetherby

West Yorkshire, LS23 7BQ

www.bl.uk

**CONTAINS
PULLOUTS**

Abstract

This thesis is a critical analysis of the use which has been made of the computer in archaeology up to the year 1972. The main chapters cover the applications in archaeology of Statistics, Information Retrieval, Graphics, Pottery Classification and Survey Reduction. A large body of Miscellaneous Applications, including Pollen Analysis, are also examined.

The majority of computer applications have been in Statistics. These applications include Numerical Taxonomy, Matrix Manipulation and Seriation, the generation of hypotheses and models, Multidimensional Scaling, Cumulative Percentage Graphs and Trend Surface Analysis. It is worthwhile to note that for small sets of data several manual methods give comparable results to complex computer analyses and at far less cost.

Computer Information Retrieval is examined in the light of its use for large bodies of specialist archaeological information, for museum cataloguing, and for the compilation of a site excavation record using a remote terminal.

The use of Computer Graphics in the production of archaeological maps, plans and diagrams is examined. Facilities include the production of dot-density plots, distribution maps, histograms, piecharts, pottery diagrams, site block diagrams with 3D rotation and perspective, sections, pit outlines and projectile point classification by Fourier analysis.

The use of the d-Mac Pencil Follower in the objective classification of pottery is described, followed by computer

analysis of the resultant multivariate data.

The use of the computer in the routine reduction of geophysical observations taken on archaeological sites is described. Complex filtering procedures for the removal of background effects and the enhancement of the archaeological anomalies are examined.

Since other workers have concentrated on the applications of statistics in archaeology, this thesis explores the relatively neglected fields of Graphics and Pottery Classification. Evidence is presented that significant advances have been made in the classification of pottery vessels and projectile points, and in the graphical output of results. A number of new programs have been developed; these include software which may be operated from a remote terminal at an archaeological site.

The P L U T A R C H System (Program Library Useful To ARCHaeologists) is described. This is a control program which uses interactive graphics and overlays to combine all the computer facilities available to the archaeologist. The individual graphics, statistics, instrument survey plotting and information retrieval techniques when combined in this way can communicate via global storage, and become even more powerful.

CHAPTER 1

THE USE OF STATISTICS IN ARCHAEOLOGY

Chapter 1

The Use of Statistics in Archaeology

1.1 It is true to say that almost all published applications of the computer in archaeology concern the use of statistics. Many of these papers are by mathematicians-turned-archaeologists or mathematicians working with archaeologist colleagues, and it is natural that the mathematical and statistical aspects of the computer should have been exploited by them. But non-numerical techniques and in particular the hardware of graphical input/output have been neglected, perhaps because mathematicians tend to regard the computer as a tool for mathematics rather than a data processor and controller of peripheral devices. In the words of the Countess of Lovelace, Charles Babbage's first computer programmer: "it can do whatever we know how to order it to perform".

1.2 The uses of statistics in archaeology may be divided into the following sub-headings:

- 1.3 Numerical Taxonomy
- 1.4 Matrix manipulation and Seriation
- 1.5 Generation of hypotheses and models
- 1.6 Multidimensional scaling
- 1.7 Cumulative percentage graphs
- 1.8 Trend surface analysis
- 1.9 Manual methods

Each of these sub-headings will be treated in turn in conjunction with the extensive bibliography. Section 1.10 contains descriptions of the statistics programs developed for this thesis, and their results.

1.3 *Numerical Taxonomy*

Numerical taxonomy as applied to archaeology concerns the attachment of numerical quantities to certain attributes of archaeological materials, whereby the description of the materials may be made more objective. By calculating suitable similarity coefficients between pairs of objects based on these numerical quantities a typology may be constructed which is based solely on the population. The subjective judgment of the archaeologist will always have a place, however, and the computer typology is intended only as a guide to him, pointing out certain apparent similarities which he may have missed.

1.3.1

The subject stands on the broad-based theory of statistics developed over the past 50 years. An important early paper by Mahalanobis (1936) gives the definition of the generalised distance coefficient between species and sub-species, based on pooled variance and covariance. The standard work *Statistical methods for research workers* by Fisher (1938) summarises most of the important statistical tests. The Fisher exact test, an exact form of the approximate χ^2 test, is defined. The uses of Normal, Poisson and Multinomial distributions are covered. The theory of goodness-of-fit (χ^2 and Fisher exact tests), regression coefficients, covariance, the analysis of variance and covariance and partial and multiple regression all have sections devoted to them, and this work is still useful despite its date of publication.

1.3.1.1

The *Normal distribution* covers all values from $-\infty$ to $+\infty$. The logarithm of the frequency at any distance d from the centre of the distribution is less than the logarithm of the frequency at the centre by a quantity proportional to d^2 . Statistical significance can only be defined in relation to a test of some hypothesis: commonly deviations exceeding twice the standard deviation are formally regarded as significant.

1.3.1.2

The *Poisson distribution* is applicable to the case where only positive integer values may be assumed by the variate. If the variate has values 0,1,2 x and the relative frequencies are

$$e^{-m} \left(1, m, \frac{m^2}{2!}, \dots, \frac{m^x}{x!}, \dots \right)$$

where the bracket extends to ∞ , then the variate is distributed in the Poisson series. The total frequency is unity since the infinite series

$$1 + m + \frac{m^2}{2!} + \dots$$

has a sum to infinity of e^m , where m is the mean occurrence of the variate in the sample of observations.

1.3.1.3

The *Multinomial distribution* gives the chance of a random sample of n containing

a_1 items of the first kind

a_2 " " " second "

a_s " " " last "

as

$$\frac{n!}{a_1! a_2! \dots a_s!} p_1^{a_1} p_2^{a_2} \dots p_s^{a_s}$$

where p_1, p_2, \dots, p_s are the probabilities of the items being in the s groups. This is a general term in the polynomial expansion

$$(p_1 + p_2 + \dots + p_s)^n$$

1.3.1.4

Goodness-of-fit may be measured by the χ^2 test or the Fisher exact test, among others.

1.3.1.4.1

$$\chi^2 = \sum \frac{(x - x_e)^2}{x_e} \quad \text{where } x_e \text{ is the expected value of } x.$$

χ^2 tables give the probability that any value of χ^2 will be exceeded for any number of degrees of freedom. $p = 0.05$ is taken as an arbitrary value, and higher values of χ^2 indicate real discrepancy. The sum of a number of χ^2 terms also falls in the χ^2 distribution.

1.3.1.4.2

The *Fisher exact test* gives an exact result, whereas χ^2 is an approximation. For fourfold (2 x 2) tables giving the percentage occurrences of the four variables in pairs, e.g.

a	b	a+b
c	d	c+d
a+c	b+d	a+b+c+d

$\chi^2 = \frac{(ad-bc)(a+b+c+d)}{(a+b)(c+d)(a+c)(b+d)}$. This is a measure of the probability of the occurrence of any set of marginal frequencies as above. The Fisher exact test gives the probability as

$$\frac{(a+b)!(c+d)!(a+c)!(b+d)!}{(a+b+c+d)!} \cdot \frac{1}{a!b!c!d!}$$

The method of conducting Fisher exact tests, however, needs care.

1.3.1.5

A *regression function* represents the mean value of one variate as a function of another, e.g. a function which represents the mean height of a person at any age is termed the regression function of height on age. Age in this case is the independent variate and height is the dependent variate. Errors in independent variates will alter the regression curve while errors in dependent variates will not in general alter the regression curve if positive and negative errors cancel. The regression function does not depend on the frequency distribution of the independent variate, but the frequency distribution of the

dependent variate will alter the regression function. A linear regression function has the form:

$$y = a + b(x - \bar{x})$$

where b is the *regression coefficient* of y on x . The dimensions of the coefficients depend on those of the variates. Several coefficients may be necessary in a regression function, e.g.

$$y = a + bx + cx^2 + dx^3$$

is a cubic. For linear regression

$$\bar{y} = a + b(\bar{x} - \bar{x}) = a$$

1.3.1.6

Covariance is defined as the expected value of the product of two variates each corrected for its mean.

1.3.1.7

Multiple regression is regression with several independent variates, e.g. the use of skull measurements to ascertain the sex of a specimen.

1.3.1.8

Partial regression is a regression of a dependent variable adjusted for the effects of a subset of the independent variables upon the remaining independent variables of a given set.

1.3.2

The detailed history of the development of the χ^2 test is given by Cochran (1952). The use of the Fisher exact test is recommended for a sample of less than 20. χ^2 is defined in this paper as

$$\chi^2 = \sum_{i=1}^k \frac{(x_i - m_i)^2}{m_i} = \sum_{i=1}^k \frac{x_i^2}{m_i} - n$$

where n is the number of items in the sample

m_i are the expected numbers in categories $i = np_i$

p_i are the probabilities of categories i

and x_i are the actual numbers in categories i . The essential point

for this definition of χ^2 is that $(x_i - m_i)^2$ has to be approximately a squared normally distributed variable. However, if m is small (say less than 5) the distribution of x is more like a Poisson variable and will be skew. Moreover, $n > 20$ does not guarantee that m is not small. Tables are available for the determination of the probability of a given value of χ^2 .

1.3.3

Completing this background of early work, Fruchter (1954) reviews the work of earlier statisticians in *factor analysis*. Spearman (1904-1950) evolved a method of matrix analysis which led to the specification of the *g-factor* (or factor expressing the general property of the variates) and the *s-factor* (or factor expressing the specific differences peculiar to variates). He explored the use of *tetrad differences* e.g. $r_{31}r_{42} - r_{32}r_{41}$ (relationships between similarity coefficients which are approximately zero in value). Holzinger (1940-1949) worked on a bi-factor theory, with the postulation of a *g-factor* and a series of non-general group factors. In cluster analysis, Holzinger and Harman defined the *B-factor* as the ratio of the average correlation of the variables in a cluster and their correlations with the variables outside the cluster. The algorithm for the adding of items to a cluster ceases to add further items when the B-factor falls to the arbitrary value of 1.3. This is an example of a discrimination criterion. Fruchter defines the *correlation matrix*, and also reviews the work of Hotelling (1933-1949) in the use of matrix calculation methods in computer programs. Computers were in their early (1st generation) days at this time, the first electronic computer (ENIAC) having been commissioned in 1946 in the U.S.A. The prototype British computers of the period were the Williams-Kilburn machine (Manchester, 1948), EDSAC (Cambridge, 1949) and its developments ACE PILOT (NPL, 1950) and DEUCE (English Electric Co.Ltd., 1952). All these were valve computers and computing techniques were at this

time in their infancy both as regards hardware and software.

1.3.4

The first uses of statistical techniques in archaeology employing a computer are thought to have been in 1959. Ihm used the Euratom IBM 650 at Ispra, Italy, to analyse a collection of Eurasian Bronze Age axes supplied by Ellisseeff, and generated a classification based on the population using numerical taxonomy (Ihm, 1961). Hiorns and Atkinson used the Ferranti Mercury at Manchester and Oxford for the discriminant analysis of 79 Wessex (Bronze Age) graves considering 21 types of grave goods (Hiorns, unpublished Diploma project in Statistics, Birmingham University, reported by Atkinson at the December 1960 Conference at the Institute of Archaeology, London).

1.3.5

The increasing availability of computers made numerical taxonomy more popular, and this increased interest is reflected in the publication of a comprehensive article in *Nature* (Sneath and Sokal, 1962) and of a textbook (Sokal and Sneath, 1963). These works define taxonomic terms, with reference to biological data. The techniques which are most important to archaeologists, and which will receive extensive treatment below are seriation, phenon diagrams (dendrograms), similarity coefficients and taxonomic distance. At this time second generation computers (employing transistors) with their increased speeds, larger storage and more advanced peripheral devices were in production.

1.3.6

Representative work on the use of multiple regression in archaeology using an IBM 7090 during this period is described by McPherron (1963).

1.3.7

A research seminar on statistics and archaeology held at the Institute of Archaeology, London on 30 May 1964 reflected the increasing interest among archaeologists. Three papers relevant to the

study of projectile points in this thesis (see paragraphs 1.10.5 and 1.10.6) were by McBurney (1964), Roe (1964a) and Bordes (1964), and concerned palaeolithic tools.

1.3.7.1

McBurney produced examples of multivariate change with time, using de Sonneville-Bordes indices for palaeolithic tools. His essential method is to observe tool frequencies and perform correlations, and he mentions the use of matrix analysis (see paragraph 1.4). He suggests that it would be useful to compare the relative efficiencies of different methods of multivariate analysis, a topic which has received more recent detailed treatment by Gower (1971).

1.3.7.2

The metrical and statistical analysis of handaxe groups by Roe has been published (Roe, 1964b). Objective measurements were taken of handaxe size, refinement and shape. Size may perhaps be linked to the available raw material, or it may be a purely functional consideration. Refinement is of cultural and chronological significance, while shape is functional and may be related to the palm size of the users. The primary measurements taken were maximum length, breadth and thickness, and weight. Histograms were plotted of the lengths and weights in a sample (for remarks on histograms see paragraphs 3.4.7 and 3.4.8). Flint is a constant enough substance for weight to be a true indicator of size, and length is also an indicator of size, having little effect on shape (these considerations have led to the similarity coefficients adopted for projectile points in this thesis). Refinement is measured as a function of thickness/breadth. To give added prominence to thin tips the ratio T_1/L may be used, where T_1 is the thickness one fifth of the way down from the tip, and L is the length. An initial indication of the profile or shape of the axe is

given by B/L , where B is the maximum breadth, but this ratio gives similar values for widely-differing shapes; a more suitable ratio is L_1/L where L_1 is the distance of the maximum breadth from the base. Finally, measurements are taken at fixed points along the length of the implement, viz. B_1 at $\frac{1}{5}$ down from the tip and B_2 at $\frac{1}{5}$ up from the base (a more advanced treatment of this principle has been employed in the similarity coefficients for pottery developed in this thesis). B_1/B_2 is a measure of the shape of the axe. A scatter diagram is constructed in which L_1/L (top-heaviness) is plotted against B_1/B_2 (symmetry coefficient) for each axe. A disadvantage of this method is that all symmetrical shapes fall at a central point (but these may be resolved using B/L). In the final presentation groups of cleavers, ovates and pointed types are separated by L_1/L and for each group B_1/B_2 is plotted against B/L . This is a simple method yielding a clustering of like shapes.

1.3.7.3

Bordes (1964) at the same research seminar calculated percentage occurrences of various types of implement and used these as the basis for cumulative percentage graphs. (q.v. section 1.7).

1.3.8

In the U.S.A. Brown and Freeman (1964) used χ^2 and regression in the study of sherd frequencies from the Carter Ranch Pueblo, Eastern Arizona, using a UNIVAC computer. It was found by these techniques that brown, textured jars show association with rooms having circular floor pits, and were probably used for cooking. Painted pottery types are associated with square fire pits and mealing bins, and were probably used in food preparation.

1.3.9

The calculation of Fourier transforms by computer became easier in 1965 with the publication of the fast Fourier transform algorithm (Cooley and Tukey, 1965). Fourier transforms have been used in this thesis for the description of projectile point profiles (see paragraph 1.10.5), a method which has not been seen elsewhere.

1.3.10

Continuing the use of numerical taxonomy in the study of stylistic change, Deetz (1965) applied the technique to Arikara ceramics using an IBM 704. Although not published until 1965, this was probably the earliest use of the computer for archaeological purposes in the U.S.A. (about 1960). Percentage matrices were produced showing the degree of correspondence between attributes of the pottery. The degree of association between pairs of attributes may be used to indicate correspondence between cultural changes and changes in ceramic styling. The custom of matrilineal residence was prevalent in Arikara settlements, and this would tend to give high stability in ceramic styling since the women, who were the potters, were living with or near their mothers, and the potting customs would be handed down. Any change in this matrilineal residence custom would lead to a diversification of styles. Thus changes in ceramic style may be taken to infer cultural changes.

1.3.11

The first of several books of collected papers on the use of computers in anthropology, archaeology and history appeared in 1965 (Hymes). A paper by Needham, a colleague of Clarke, concerned computer methods for classification and grouping. The paper explains the need for a coefficient of overall resemblance, but points out that if certain properties are weighted in a similarity coefficient then one

must be extremely clear about the reasons for so doing. Subjective weighting may destroy the very impartiality which resource to computers could supply. The computer cannot replace the insight of the scholar, but it may indicate classifications which were not suspected previously, and it may also reveal the lack of real basis for an accepted classification. Needham uses the technique of not recording zero or very small coefficients of similarity. He defines a *clump* (more usually called a *cluster*) as a subset of elements such that if any foreign element is moved into it, or any included element moved out of it, the total number of significant links crossing the boundary of the subset is increased. He also defines an algorithm to isolate a *clump*: divide the set of elements arbitrarily into two; examine each element to see whether by transferring it to the other side of the boundary the number of links are decreased; take precautions to ensure that not all elements migrate to one side of the boundary; at minimum linkage a *clump* has been isolated.

1.3.12

In 1965 also appeared the first of a series of papers concerning clustering methods which aroused a controversy which has continued to the present. The protagonists are a Cambridge group (Needham, Jardine, Jardine and Sibson) and a group from Australia (Lance, Williams, Clifford and Dale), and the controversy concerns clustering methods, automatic classification, taxonomic hierarchies and the rigour of the associated mathematics (Lance and Williams, 1965, 1966, 1967a, 1967b, 1967c, 1968, 1971; Williams, Clifford and Lance, 1971; Williams, Lance, Dale and Clifford, 1971; Jardine, Jardine and Sibson, 1967; Jardine and Sibson, 1968, 1971; Sibson, 1971a).

1.3.12.1

Lance and Williams (1965) describe a *monothetic*

method of subdividing a population of individuals specified by binary attributes. This is similar to the *Roth technique* used for the selection of an efficient set of tests in diagnostic programming, with the main difference that χ^2 is used as the splitting criterion instead of the (number of 1s) x (number of 0s) product. The data is specified as: number of individuals; number of attributes; normal or inverted file; data coded in the form 1/3/5-8/10 for attributes 1,3,5-8 and 10 equal to 1, or octal form 5364 for the same attributes. A mask card is used to determine which attributes are to be analysed.

1.3.12.2

Lance and Williams (1966) define *cluster analysis* to be *non-hierarchical*. *Hierarchical* methods seek to find the most efficient step at each stage in the progressive subdivision of a population. In practice, the user must decide either to optimise the clusters (non-hierarchical) or the route (hierarchical). Hierarchical methods assign items irrevocably, but clustering methods allow the migration of items. Hierarchical classifications have computational advantages over cluster analyses. Some flexible CDC 3600 programs are described with nearest neighbour or centroid sorting strategies, and with four different similarity coefficients. The *centroid sorting* and "*information statistic*" or "*non-metric*" coefficients appear to give the best results.

1.3.12.3

Jardine, Jardine and Sibson (1967) in their first paper of the series criticise the *average-link* clustering method, claiming that it is not sufficiently rigorous from a mathematical point of view. The average-link method uses the strategy that a new item joins an existing cluster if its similarity to the mean properties of the items in the cluster is above a defined discrimination level. The only

rigorous method is *single-link* cluster analysis, which uses the strategy that a new item joins an existing cluster if its similarity to *any* item in the cluster is above a defined discrimination level. Despite Jardine, Jardine and Sibson's misgivings, the average-link method gives much more useful results in archaeological applications than the single-link method. Moreover, the single-link method has notorious chaining properties which cause quite diverse items to be linked, which is most undesirable in an archaeological classification.

1.3.12.4

Lance and Williams (1967) propound a general theory of classificatory sorting strategies, dividing the methods into hierarchical systems and clustering systems. In hierarchical systems they define five conventional strategies: *nearest neighbour* (single-link) - in this chaining is a severe disadvantage; *furthest neighbour* (single-link); *centroid* (average-link) - similarity measures used are squared Euclidian distance, correlation coefficient or non-metric distance (part Boolean, part numerical, usual for mixed data) but a disadvantage is that this is a weighted strategy and the characteristics of a minority group are virtually lost; *median* - removes the weighting effect of the centroid strategy, the new group being sited at the mid-point of the shortest side of the triangle with vertices at the two groups to be joined and a third reference group; and *group average* (another type of average-link method). These strategies have three major properties: *combinatorial/non-combinatorial*, i.e. the computer may or may not derive next values of the coefficients from existing values; *compatible/incompatible*, i.e. measures calculated late in the analysis are or are not of exactly the same kind as those calculated early in the analysis; and

space-conserving/space distorting, i.e. the space properties remain the same during the analysis, or the model behaves as though the space in the immediate vicinity of the group has been contracted or dilated. In *space-contracting* strategies a group will appear on formation to be nearer to some other items than before; the probability of the incorporation of these items has increased and chaining eventually results. In *space-dilating* strategies groups appear to recede on formation; the items not yet allocated to groups are thus more likely to form nuclei of new groups, leading to the formation of groups of non-conformist peripheral items. Lance and Williams next design a "flexible" strategy with the distance function

$$d_{hk} = \alpha_i d_{hi} + \alpha_j d_{hj} + \beta d_{ij} + \gamma |d_{hi} - d_{hj}|$$

where d_{hk} is the distance between a group k (newly-formed from groups i and j) and another item h . The coefficients α_i , α_j , β and γ can be related and in particular if

$$\alpha_i + \alpha_j + \beta = 1, \alpha_i = \alpha_j, \beta < 1 \text{ and } \gamma = 0$$

the flexibility of the strategy lies in its space-distorting properties. As β approaches +1, it is increasingly probable that the apparent distance from the first group to the nearest item will always be less than any remaining item to item distance, hence chaining results because of space-contraction. As β falls through 0 to negative values the strategy becomes space-dilating and groups become increasingly non-conformist. The best value for β seems to be a small negative fraction, when the strategy seems to be *space-conserving*.

Comparing the flexible strategy with conventional strategies, it is suggested that correlation coefficient rather than Euclidian distance is the best measure if "shape" is more important than "size". In the centroid strategy the Euclidian distance measure groups items only

weakly, but use of the flexible strategy enables any degree of grouping to be obtained.

The treatment of multi-level data is also covered in a note by Lance and Williams. They suggest four methods: a) All m_j levels of a single item may be regarded as independent - there is little to be gained from this. b) The m_j levels may be averaged - this discards important information about fine irregularities. c) All m_j pairs of readings for two items may be compared; this assumes that j is constant for all items, and that the k^{th} level of any item corresponds in some way to the k^{th} level of all other items; a dissimilarity measure of the form

$$\frac{1}{m_j} \sum_{k=1}^{m_j} \frac{|x_{1k} - x_{2k}|}{(x_{1k} + x_{2k})}$$

is suggested where x_{nk} is a positive number representing the value of the k^{th} level of the n^{th} item. This lies in the range 0 to 1 and is dimensionless. Missing values of x are catered for by omitting the calculation for the imperfect levels and reducing m_j accordingly. A method of treating pot profiles as multi-level data has been developed in this thesis (see paragraph 4.4.3). All pairs of readings are compared in this method, but Euclidian distance is used as a measure of dissimilarity. d) The fourth method of treating multi-level data suggested by Lance and Williams is to fit curves for each item and compare the parameters. For this they suggest that the computation would be formidable. Nevertheless, a similar method to this has been developed in this thesis for the comparison of projectile points (see paragraphs 1.10.5 and 1.10.6). The curve fitting process employed is the Fourier transform, and the parameters of the first five orders are compared, using Euclidian distance as a measure of dissimilarity and weighting each order by the inverse of the power series so that the

first order does not swamp the significant differences for the higher orders. The computation time is not found to be formidable if only a limited number of orders are computed.

Lance and Williams contrast clustering systems with hierarchical systems. Agglomerative hierarchical strategies suffer from the "*migration problem*", i.e. items classified early on in the analysis may in the final classification be placed in another group. Clustering methods should allow migration at all stages. The processes in a clustering method are: initiation of clusters; allocation of new items to existing clusters and/or fusing and splitting of existing clusters; determination of the profitability of further allocation; and a *migration* facility. The *k-means* clustering method seems to be the most flexible. k is the number of clusters sought. A group mean statistic is necessary, usually based on Euclidian distance. One way of commencing the procedure is to divide the population of n items arbitrarily into k groups of n/k items. Another is to generate k points in space to which items are allocated. In both cases group means are calculated for all groups. Each item is then examined in turn and may migrate to another group if it is more similar to the group mean of that group than to the mean of its own group. If two groups come closer in mean than a defined value they are fused, and k is reduced. A group may also be split into two if its variance is too high, and k is increased. An upper limit to allocation is also chosen, and if an item cannot be added to its most suitable group it forms a new group and k is increased. Lance and Williams suggest that hierarchical systems should be modified to include a migration facility; and that for clustering systems a random subset should first be classified in a hierarchical system

and the means of the groups so obtained used as nuclei for the k-means clustering algorithm.

1.3.12.5

Jardine, N. and Sibson (1968) in their reply to Lance and Williams' papers emphasize that the *single-link* cluster method is the only commonly-used method acceptable to mathematicians applying rigorous definitions. The other commonly-used methods, *centroid*, *median*, *average-link* and the *flexible* method of Lance and Williams are all discontinuous. The defect of single-link is, however, chaining. The remedy is to consider non-hierarchic (overlapping) classificatory systems. The *k-partition* is defined as a partition which allows a maximum of $k-1$ objects in the overlaps between the classes that belong to it. The system is hierarchic if $k=1$, overlapping if $k > 1$. A sequence of clustering methods: single-link, double-link, triple-link etc. may be defined. These may be illustrated on tree-diagrams, which are graphs with nodes and edges. There is often severe tangling of the edges of these graphs, and to prevent this it is best to perform non-metric multidimensional scaling which is best represented on a horseshoe-shaped 2D diagram (see Kendall, 1970). An improved algorithm based on the *k-partition* clustering method has been designed by Cole and Wishart (1970).

1.3.12.6

Lance and Williams (1968) define a new divisive monothetic method for the classification of data specified by binary attributes, the *information statistic*. This has the advantage of computational speed and rigorous stopping rules. The algorithm has been programmed for the CDC 3600.

1.3.12.7

Sibson (1971a) presents criteria which he considers essential for any general clustering algorithm and shows that the *flexible*

algorithm of Lance and Williams fails to satisfy these criteria. The difference between *methods* and *algorithms* is suggested by Sibson to be the root cause of the confusion. He reiterates the statement that the single-link method is the only well-defined method of those mentioned by Lance and Williams.

1.3.12.8

Jardine, N. and Sibson (1971) take a series of "pragmatic" criteria for the satisfactory construction of a classificatory system put forward by Lance and Williams, and either agree with or provide convincing arguments against them. It is concluded that the controversy can be partially resolved by consideration of the reasons for classification. Constraints on homogeneity of clusters or on size or number of clusters are incompatible with the Cambridge criteria of "adequacy" for hierarchic clusters. It is therefore suggested that non-hierarchic methods should be used in cases where hypotheses are to be generated or confirmed.

1.3.12.9

Lance and Williams (1971) develop the *information statistic* to be applicable to mixed data (Boolean and numerical).

1.3.12.10

Williams, Clifford and Lance (1971) reiterate their previous arguments, discounting the "mathematical criteria of classificatory excellence" of Jardine and Sibson. The single-link method, the only one accepted by Jardine and Sibson, is stated to be quite unable to provide the powerful clustering required for some applications.

1.3.12.11

In the last paper of the controversy to have appeared, Williams, Lance, Dale and Clifford (1971) explain the differences in linguistic usage between the Cambridge and Australian schools, and agree that

the Australian methods do not meet the Cambridge criteria. The Australian school considers that there exists a further set of criteria which in certain circumstances should over-ride. Cambridge ignores those strategies which need the raw data matrix rather than the primary dissimilarity matrix. The Australian school does not agree with the distinction between *methods* and *algorithms* put forward by Cambridge, and the word *cluster* has many different meanings. Counter arguments are produced for all the Cambridge criteria not met by the Australian methods. The "pragmatic" criteria are reiterated as:

- a) the grouping must be more intense than the original dissimilarity measures, i.e. must produce a small number of clearly-defined groups.
- b) the grouping must be relatively insensitive to aberrant items and inaccuracies in data.
- c) the linkage criterion should not necessarily, or even usually be invariant over the entire population, in order to accommodate recognised groups which are much more loosely linked than others. An adaptive system is necessary.

The opinion reached in this study is that much of this controversy is caused by different linguistic usage, but in the remaining differences probably both schools have something to offer. It seems pointless to adhere blindly to rigorous mathematical arguments when it has been proved that the average-link method gives sensible and useful results in many archaeological applications. On the other hand, the proliferation of methods without adequate theoretical background leaves the archaeologist in some doubt as to which he should use, and perhaps leads to such subjective choice as the automatic classification methods were originally constructed to avoid.

1.3.13

Binford and Binford (1966) carried out an analysis of Mousterian artefacts of the Levallois facies. Five clusters were found for two representative sites (Shelter I, Jabrud, Syria and Mugharet-es-Shubbabiq, a cave site near Tiberias, Israel): (a) borers etc., (b) a hunting and butchering kit, (c) Levallois flakes, (d) denticulates and (e) points. Variations through time were found in the proportions of these factors. This is a representative example of the use of *factor analysis* for the discovery of archaeological types.

1.3.14

On a more theoretical note, Gower (1966) investigated some distance properties of latent root and vector methods used in multi-variate analysis. This work was prompted by dissatisfaction with the many reported applications of factor analysis of matrices. The interpretation of these methods can be better understood by examining the distances between the items. A common method is the *Q method* (analysis of a square matrix representing degrees of association between artefacts) which finds a set of points Q_i such that the distance d_{ij} between two points Q_i and Q_j is given by

$$d_{ij}^2 = a_{ii} + a_{jj} - 2a_{ij}$$

where a_{ij} is the number of times an item of type i is associated with an item of type j

a_{ii} is the total number of items of type i

a_{jj} is the total number of items of type j

It is found that this is a desirable property for many coefficients of association. The method of *principal components analysis* is used to find the best fit of the data into a smaller number of dimensions.

The *R method* analyses a square matrix representing degrees of association between attributes, and the Q and R methods are duals, i.e. both give rise to the same set of points (in terms of interpoint distances) in multidimensional space. The above method of operation on a Q matrix is also the dual of principal components analysis operating on an R matrix. Gower also examines Sokal's taxonomic distance and Mahalanobis' D^2 statistic in the same light. The Kruskal and Shepard method of multidimensional scaling also receives attention, but in this case the interest lies in the use of rank order of similarity rather than the actual distances.

1.3.15

Moberg (1966) reports an experiment for the determination of technical traits of assemblages of flint blades and flakes. In this study particular attention is paid to the problem of broken flakes and blades.

1.3.16

Rowlett (1966) reports the availability of a FORTRAN program for the determination of which attributes of spearheads will be most useful for chronological studies and which are useful for cultural group attribution. The scope of the study covers iron spearheads from Hallstatt I to La Tène IIb in age, using generalised distance.

1.3.17

A useful textbook on modern factor analysis has been published by Harman (1967). This is a good introduction to the terminology, concepts and methodology of factor analysis. There are detailed references to computer software. The numerical processes of factor analysis, the solution of linear equations, matrix inversion and rotation of axes are also covered.

1.3.18

With the advent of 1968 third generation computers became

generally available, with a consequent increase in speed of arithmetic and input/output processes. The number of reported applications of the computer to archaeology also increase. Cowgill (1968) discusses the application, advantages and limitations of multidimensional scaling, factor and cluster analysis. He discusses representative applications, e.g. La Tène brooches (Hodson), Mousterian assemblages (Binford, see paragraph 1.3.13), pottery from Carter Ranch Pueblo (Brown & Freeman, see paragraph 1.3.8) and Teotihuacan (Cowgill). The dangers of using correlations based on inadequate samples are stressed, and one necessary condition for sample adequacy is suggested. It is argued that multidimensional scaling probably gives better results than seriation.

1.3.19

In a composite work on statistical geography edited by Berry and Marble (1968), Harbaugh and Preston discuss the application of Fourier series to spatial problems. The main use of this in archaeology is for trend surface analysis (q.v. section 1.8), but in this thesis the Fourier transform has been applied to the analysis of the profiles of projectile points (see paragraphs 1.3.12.4 and 1.10.5). The use of the trapezoidal rule for the numerical integration of empirical data multiplied by sine or cosine is covered. The use of the power series (sum of the squares of the sine and cosine coefficients for a harmonic) is described. The power series has been used in the projectile point analysis of this thesis to weight the Euclidian distances between corresponding coefficients of two items, thus preventing the swamping of significant differences in higher order coefficients by the dominant first order terms.

1.3.20

Rhoads (1968) reports ALGOL software for the IBM System/360

series of computers fitted with magnetic tapes, discs and a digital incremental plotter. The purpose of the programs is to analyse archaeological data using statistical techniques and to perform graphical output.

1.3.21

Roe (1968) has published a more detailed classification of British Lower and Middle Palaeolithic handaxes. This is a development of the work reported in paragraph 1.3.7.2

1.3.22

Sounding a note of caution, Brothwell (1969) remarks on the possible misuse of statistics and computers in archaeology. He makes the point that the computer can indicate the distances between cultural assemblages but cannot indicate cultural significance.

1.3.23

Hodson (1969) in two similar papers describes the applications of statistics to a collection of bronze brooches of the pre-Roman Iron Age from a linear cemetery at Münsingen-Rain, to Upper Palaeolithic assemblages, and to European copper and bronze objects. The methods used are *average-link cluster analysis* with *dendrogram* (phenon diagram) output, using Euclidian distance as the measure of association; *principal components analysis*, in which the first axis generated accounts for the maximum possible variance in the sample, each subsequent axis is independent of all others, and the end result is an n-dimensional diagram using the generated axes; and *multidimensional scaling* (Kruskal). The results are compared with Bordes' intuitive results for Palaeolithic assemblages, and with archaeologists' intuitive classifications of the Münsingen-Rain brooches. The formula used for Euclidian distance applied to metal

analyses is

$$d_{ij} = \sqrt{\frac{\sum_k (x_i^k - x_j^k)^2}{\sum_k (x_i^k)^2 + \sum_k (x_j^k)^2}}$$

where x_i^k is the percentage of the k^{th} metallic element in the i^{th} object. A similar formulae to this has been used in the projectile point profile analysis of this thesis (see paragraph 1.10.6). The Euclidian distance above was incorporated in a program written for the Cambridge University Titan machine by Hartzig. Hodson's average-link cluster analysis program was written by Hirsh and Hirsh. The results are analysed into dendrogram form but *not* plotted by computer, but instead drawn out by hand from the results of the computation (for the automatic features of dendrogram plotting by programs developed for this thesis see paragraph 3.4.15). Hodson comments that the *average link* clustering method seems to give satisfactory results for the archaeological classification topics he has studied. Hodson briefly mentions the use of +, - and/to signify presence, absence and inapplicability. It is possible that these ideas may be extended using the 1, 0 and d (don't care) of Boolean logic.

1.3.24

Thomas (1969) reports FORTRAN IV software written for the IBM 9044 computer with digital incremental plotter for the output of dendrograms. The aim is to investigate the application of numerical taxonomy to archaeology. Examples studied are petroglyphs, projectile points, entire sites and assorted tool types. The main application is a settlement pattern study in Central Nevada. The method is to construct a matrix of correlation coefficients based on over 50 attributes. Clustering is performed and dendrograms plotted. These separate runs are then combined into larger dendrograms to form a nested classification.

1.3.25

Cole and Wishart (1970) describe an improved algorithm based on the *k-partition* clustering method of Jardine, N. and Sibson (1968) reported in paragraph 1.3.12.5. The algorithm generates overlapping clusters and makes a significant saving in computer time over the original algorithm.

1.3.26

Collins (1970) investigates stone artefact analysis and the recognition of culture traditions. He is doubtful whether the application of computers to the problem of the classification of stone artefacts is appropriate, because of the lack of "intuitive" insight about the mixed nature of some classifications, or the possibility that two classifications are different stages of a common tradition. Computer programs have also assumed that Bordes' 63 tool types are of equal significance, which Collins states is clearly not so. He concludes that "computers will only be as valuable as the ideas which they are testing are sophisticated".

1.3.27

The use of *canonical variates* has been studied by Graham (1970), applied to the discrimination of British Lower and Middle Palaeolithic handaxe groups. Canonical variates are linear combinations of the form

$$\sum_{i=1}^n \left\{ (\text{weighting factor}) \times (\text{measure of variable } i) \right\}$$

The purpose of canonical analysis is to determine the weighting factors which give the best possible discrimination between assemblages of artefacts. The procedure is to take measurements and ratios for all items, then to apply the χ^2 test of significance to find significant canonical variates. The analysis was carried out on data supplied by Roe. A 2D diagram is formed by plotting values for the two most

significant canonical variates on cartesian axes. The result could be submitted to cluster analysis, but human pattern recognition is found to be superior in this case. The derived groups were submitted to Roe for his comments, and he sees no great disagreement between Graham's results and his own, apart from one site.

1.3.28

A handbook of multivariate methods programmed in ATLAS Autocode by Hope (1969) has been reviewed by Hiorns (1970). Algorithms are described for the analysis of variance and covariance, principal components analysis, regression and canonical analysis. This handbook is an adequate description of a worthwhile set of programs.

1.3.29

Hodson (1970) again reviews the main techniques of numerical taxonomy applied to archaeology. *Multidimensional scaling* has now replaced the Brainerd-Robinson technique of ordering shaded matrices, which is not a precise method of cluster-analysis. *Single-link* cluster analysis has been shown by Jardine and Sibson (see paragraph 1.3.12) to be the only method regularly in use which possesses the formal properties required by their comprehensive theory of clustering. *Average-link* cluster analysis suffers from the disadvantage that items may become trapped at an early stage of the analysis in clusters from which they cannot later escape. Thus some similar items may not join the same cluster until several other linkages have taken place. This is illustrated by a rather unsuccessful dendrogram produced on the line printer. *Double-link* cluster analysis was evolved by Jardine and Sibson (1968). In this strategy those items which would immediately be fused by the single-link strategy do not join the cluster but are allowed to overlap. A second link with another item in the cluster does cause fusion. *Triple-link* cluster analysis requires three links

for fusion and allows two overlaps. The idea may be extended to define a whole series of clustering strategies. The use of *Venn diagrams*, a technique borrowed from logic, to illustrate the overlap of clusters is described, and it is also possible to use *dendrograms* with several links per item, although this becomes rather complex. The undesirable chaining properties of single-linkage are still evident in double-linkage. The *k-means* clustering strategy seems most promising. The *k-means* are averages for hypothetical cluster centres. Items are compared with these centres, which may become progressively modified as a result, but a complete set (half-matrix) of pair-relationships need not be retained, an important consideration when a large number of items have to be investigated and computer storage is limited. The measure of homogeneity used for these clusters is the *sum-squared error*, or sum of the squared Euclidian distances between constituent items and the cluster centres. The algorithm for *k-means* is as follows:

- a) each unit is assigned to the current cluster centre to which it is closest;
- b) new averages are calculated for each current cluster;
- c) information about the clusters is printed out;
- d) the cluster (or clusters) with the highest sum-squared error is split by finding the variable with the highest standard deviation, and dividing the cluster into two according to their score on this variable; *k* is increased as necessary;
- e) the pair (or pairs) of clusters that when joined minimally increase the sum-squared error are joined; *k* is reduced as necessary; the cycle then repeats from a), and continues until no further significant improvement is obtained, or the rate of improvement is too slow, or a specified number of cycles have been completed.

The progress of the k-means algorithm may be represented as a dynamic dendrogram. Items may migrate between clusters, and this is an improvement on the average-link method. Two other methods are described, *principal components analysis* and *multidimensional scaling*. A principal components plot is compared with a scalogram (multidimensional scaling output) of cluster centres from a k-means analysis, and it is evident that these show similarities.

1.3.30

Mellars (1970) makes some comments on the concept of "functional variability" in stone-tool assemblages. While appreciating the methodology of such authors as Binford and Freeman, Mellars questions the conclusions of "functional variability", preferring a concept of differing traditions. He criticises Binford's 2D distributions based on factor analysis as being "vague and diffuse" and requiring a lot of subjective judgment in their interpretation.

1.3.31

A FORTRAN algorithm which produces an unstratified hierarchy of clusters is published by van Rijsbergen (1970), designed in collaboration with Jardine and Jardine.

1.3.32

Gower (1971) continues with analysis of different multivariate analyses of the same data reported in paragraph 1.3.14. He investigates the problem of mapping n points with Euclidian distances such that the distances bear some relationship to the corresponding dissimilarity coefficients, and the construction of a dendrogram. Examples given are the analysis of the skulls of modern *Homo sapiens*, Upper Palaeolithic *Homo sapiens*, Middle East Neanderthal, European Würm Neanderthal, *Homo erectus* and *Australopithecus africanus*. Mahalanobis distances are computed for the $6(6 - 1)/2 = 15$ population differences,

for eight measurements. The matrix corresponding to each measurement is rotated to give the best fit with the other matrices, and the results collated to give a single matrix. This is a solution to the problem of combining multivariate data to give a single similarity coefficient.

1.3.33

Hodson (1971) describes some developments in *k-means* cluster analysis. The method is essentially the same as that reported in his 1970 paper (see paragraph 1.3.29), with the difference that some features have been incorporated to avoid the detection of local rather than global maxima, a problem encountered in all hill-climbing algorithms. The splitting operation is performed and the clustering criterion is checked for 2, 3 up to the maximum number of clusters. When the maximum number of clusters is reached, the fusing operation is performed, the new clusters checked against the best performance for the number of clusters, and if this represents an improvement, the splitting operation is entered and followed until no improvement on previous results is achieved. At this point the best result for that level is substituted and the fusion resumes. The procedure stops when a two-cluster result which does not represent an improvement on a previous two-cluster is obtained. A complete spectrum from two up to the maximum number of clusters is then available to choose from, with values of the clustering criterion. Mahalanobis rather than Euclidian distance is preferred for the clustering criterion calculated from the cluster and global centres. The improved method has been tested on Roe's handaxe data, and seems to suggest three stylised types of handaxe. The data consists of a randomly-chosen sample of Roe's measurements (see paragraph 1.3.7.2), and it will be interesting to see whether the profile similarity coefficients

developed in this thesis for projectile points, equally applicable to handaxes, give similar results (see paragraphs 1.10.5 and 1.10.6).

1.3.34

Sibson (1971) in a short paper describes some computational methods in cluster analysis. A dissimilarity coefficient is defined as zero between an object and itself, and otherwise non-negative. The coefficients are commutative i.e. $d(x,y) = d(y,x)$. Dissimilarity coefficients may be obtained from most similarity coefficients by subtraction from a fixed number. The $n(n-1)/2$ dissimilarity coefficients of a half-matrix less diagonal may be read in the order 2-1; 3-1, 3-2; 4-1, 4-2, 4-3; 5-1 $n-(n-1)$. Algorithms for performing cluster analyses on such data fall into four groups. Group 1 deal with the coefficients value by value, and the values may then be discarded. *Single-link* cluster analysis is such a method. Group 4 require access to the entire half-matrix throughout the operation. *Average-link* cluster analysis is of this type. Groups 2 and 3 are intermediate between these two extremes; Group 2 algorithms read the coefficients a part-row at a time, while Group 3 algorithms require some specified number of scans of the entire half-matrix but not complete random access. Sibson reiterates his preference for the single-link method because its mathematical properties are known, and his dislike of the average-link method.

1.3.35

Borillo (1972) reports FORTRAN IV software written for the IBM 360 and UNIVAC 1108 for the establishment of classifications for incompletely-specified objects. This is a neglected field, which is strange since archaeological data is often incomplete.

1.3.36

FORTAN V programs for the UNIVAC 1108 are described by de la Vega (1972). The card input consists of numerical values of the

attributes of objects, and the programs construct and print hierarchical classifications using average-link or single-link methods.

1.4 *Matrix Manipulation and Seriation*

The storage of archaeological data in matrix or half-matrix form in the computer has been described above. Some of the earliest computer programs applied to archaeology manipulated these matrices, with the aim of isolating significant groupings. It is more usual to manipulate not the incidence matrix (in which, for example, rows may represent specific artefacts and columns attributes of the artefacts) but one of the two possible square matrices formed by the matrix product of the matrix with its transform, and these are known as the Q and R techniques. Another type of matrix may represent assemblages of artefacts from different time periods, e.g. graves or whole sites. Using the basic assumption that artefact styles are invented, grow to popularity, then decline, it is possible to re-order the matrix so that the best sequence of this type is obtained (this may be done by calculating similarity coefficients between all the assemblages, and arranging for all the highest similarity coefficients in the reordered matrix to lie near the diagonal). The method is known as the chronological ordering of artefact assemblages, or *seriation*.

1.4.1

The essential theory behind the seriation method was first published in 1951 by Brainerd and Robinson. This is a classic example of the cooperation of an archaeologist and a mathematician to produce a method which neither could have produced alone. The basic ideas behind the method are described, and illustrated by the ordering of suitable matrices. Lehmer objects to the method on the grounds that Robinson's coefficient of agreement takes no account of differences in the sizes of the assemblages, and that large assemblages are necessarily more representative of their environment than are small collections. Lehmer proposes to compensate for sample size by using the complement of the mean standard error of the differences which make up Robinson's coefficient of agreement. Robinson and Brainerd in their reply

(1952/53), however, state that Lehmer confuses the problem of estimating a population parameter with the problem of testing the statistical significance of an observed result. The method is shown to be correct even when sample sizes differ widely. All assemblages give estimates of the proportions in the total population, regardless of their actual size. Lehmer's method does give an indication whether the difference between the proportions in two assemblages is statistically significant, but gives a biased estimate of the population difference of the two assemblages. Hypothetical data with widely-differing sample sizes is used to prove the superiority of Robinson's technique over Lehmer's method.

1.4.2

Belous (1953) applies the new seriation ideas to the Central Californian chronological sequence. The sequence predicted by the seriation method is checked against the known stratigraphy, superposition of buildings of different archaeological periods and other field observations.

1.4.3

No further new ideas about seriation appear until 1959, when Meighan and Ascher in separate papers describe a new graphical technique using three types of artefacts. The occurrence of each type of artefact is expressed as a percentage of the total population of the three artefacts, and the three percentages are plotted on *three-pole* graph paper, the axes of which form an equilateral triangle. When each assemblage has been plotted as a point on this diagram, the best-fit straight line is drawn through the points. Dropping perpendiculars to the line from the points, the ordering of the points then indicates a possible chronological order for the assemblages. The mathematical justification behind this method is given by Ascher in the companion paper. He gives an algebraic explanation and suggests a simplification which makes it

possible to plot the distribution on cartesian rather than three-pole graph paper. Since the sum of the three artefact occurrences is 100%, the percentage may be treated as a, b and (100-a-b) giving only two independent variables. Again, each assemblage is plotted as a point on the cartesian graph paper, using axes a and b, and the best-fit straight line is drawn through the points. Perpendiculars to the line may then indicate the chronological ordering of the assemblages.

1.4.4

Moberg (1961) comments on the controversy which the new seriation methods had caused among archaeologists. He suggests that the quantitative procedures used in ordering assemblages replace similar methods used intuitively by archaeologists, and that the new methods borrowed from the natural sciences do not introduce radically new ideas.

1.4.5

The standard statistics textbook by Hoel (1962) gives an introduction to the statistical technique of estimation by *maximum likelihood*. It is suggested by Doran (1970) that this method can be used to express the likelihood L_i of obtaining an observation E_i as a function of E_i and unknown dates for archaeological phases of a site:

$$L_i = \text{LIKE}_i (E_i, D_1, D_2, \dots, D_n).$$

If this is done, then the task of finding the most likely chronology for the site is one of maximising the function

$$H = L_1 \times L_2 \times \dots \times L_m$$

over m observations. However, this is not always easy, since there are frequently too many unknowns leading to likelihood surfaces with saddle points and discontinuities.

1.4.6

The first computer program for the ordering of matrices after the method of Brainerd and Robinson appears in 1963 (Ascher and Ascher). The algorithm holds aside rows of the matrix which are not "well-behaved" i.e. which give inconsistent results in the positioning routine.

The resulting order of the well-behaved rows, viz. those which can be ordered so that the coefficients increase towards the central diagonal and then decrease, is output. In the analysis each new row is tried as row 1 and column 1, then as row 2 and column 2, and so on until its best position is found. The algorithm is coded in ALGOL for the Burroughs 220 machine, and has been improved by Kuzara, Mead and Dixon (1966, q.v. paragraph 1.4.12).

1.4.7

Dempsey and Baumhoff (1963) develop the seriation method by suggesting that presence/absence counts for artefacts may be used. The method arranges assemblages in a sequence so that, as far as possible, like is adjacent to like and far from unlike. This procedure does not indicate any degree of relative spacing in time, e.g.

A BC D E appears as

A B C D E.

The method has been criticised by Lipe (1964) for two specific situations:

- a) when artefact mixture has occurred
- and b) when the types of artefact are based on continuously varying attributes and overlap one another to some extent because of normal variation in the attributes.

Lipe suggests that the Dempsey and Baumhoff technique may not be as useful as the Brainerd-Robinson method in these situations. Dempsey and Baumhoff object to Robinson's use of relative frequencies which causes types to be weighed in proportion to their abundance. They argue that types which occur with low frequency may be among the best time indicators. But Lipe indicates that mixture of artefacts from different periods, a common occurrence on an archaeological site, will lead to invalidation of Dempsey and Baumhoff's method, since nearly every type will occur in every time unit. Robinson's method, on the

other hand, will not be affected much. In the second type of situation (continuously varying attributes), a change from long to short projectile points would not be detected by Dempsey and Baumhoff's method, since the statistical uncertainty leads to both types being present in all assemblages, obscuring a real cultural change from the use of long to the use of short projectile points. It seems that both methods have uses in specific situations, and may be regarded as complementary.

1.4.8

One of the earliest applications of matrix typology in Britain was by Clarke (1963), to British Beaker pottery. The 1963 paper is a theoretical study, the actual results of the British Beaker pottery analysis being reported elsewhere (see paragraph 4.3.2). Clarke compares the results from an *uncompensated matrix* (expressing the number of times each attribute occurs with every other attribute), a method which though statistically inelegant has the advantages of directness and speed, with a *compensated matrix* using Tanimoto's compensating factor, and with cluster analysis, and finds that "agreement is impressive". It is clear, however, that these methods are imperfectly understood, and the work of Clarke has been criticised by Matthews (1963). Matthews describes the Q and R techniques, which analyse square matrices (artefact-artefact and attribute-attribute respectively). She demonstrates that two dissimilar assemblages of artefacts may give rise to identical similarity matrices of the R type. The use of the R technique is therefore unreliable for the identification of types, and this calls into question Clarke's results for British Beaker pottery. There is a need for a method to describe the shape of a pottery vessel, which cannot be defined easily by a single equation or factor. This thesis develops such a method (see paragraph 4.4.3).

1.4.9

Deetz and Dethlefsen (1965) consider the effect of the movement of cultures upon the seriation method. Different cultural influences may

reach different areas at quite different times; two assemblages A & B may be contemporary but A may be interpreted as later than B because certain attributes characteristic of A, the more developed culture, have not yet had time to spread to B because of geographical distance. This phenomenon is analogous to the Doppler effect in physical science, where the observed frequency and wavelength depend on the velocity of the transmitter. Thus purely spatial variations might be interpreted as chronological variations by the seriation method, and if areas of different sizes and geographical location are sampled quite different results might be obtained.

1.4.10

There has also been some work on the seriation method in Russia (e.g. see *Kamenetskiĭ*, 1965). This paper uses percentages of pottery types in the seriation method, and there are no innovations over work in U.S.A. and Europe.

1.4.11

Tugby (1965) reviews methods of matrix analysis together with cluster analysis and factor analysis, concluding that matrix analysis of series of items rather than types seems to give best results in archaeology (cf. Matthews, 1963 see paragraph 1.4.8).

1.4.12

Ascher (1966) reports continued work on seriation methods, concentrating on mathematical analysis of the problem, beyond that described above (paragraph 1.4.6). The algorithm of Ascher and Ascher has been improved by Kuzara, Mead and Dixon (1966). The advantages of the improved algorithm include working with the whole matrix from start to finish instead of with one row and one column at a time, and more exhaustive row-column comparisons rather than the permanent placing of each row and column as soon as it satisfies the criteria. The program has provisions for specifying different ordering criteria and/or allowing consistent use of a chosen criterion for evaluating different matrix

configurations. The program also avoids built-in input order bias.

1.4.13

Sackett (1966) describes measurement techniques for stone tools (length, butt angle, front contour, etc) and the use of these measurements in seriation methods. No specific computer work is described but several references to computer work (Binford, Clarke, Kruskal, Spaulding, Tugby, Brainerd and Robinson) are cited. This thesis develops an integrated system of stone tool measurements using the d-Mac pencil follower (see paragraphs 1.10.5 and 1.10.6).

1.4.14

Other notable exponents of the seriation method in the U.S.A. are Hole and Shaw (Hole, 1966; Hole and Shaw, 1967, 2 references). They experiment with different methods of automatic determination of order from a matrix. The Meighan 3-pole plot is also used (see paragraph 1.4.3). Examples are given for Iranian artefacts from the Mousterian onwards (pottery, flint tools and miscellaneous artefacts). Software is written in ALGOL for the CDC G-21 machine. The methods used for output of ordered matrices are line printer and line-drawing display unit. This work is reviewed by Cowgill (1968) in an article which is far more than just a book review; it is rather a critique and summary of all the work to date in seriation, and a clear presentation of the values and limitations of the technique.

1.4.15

A paper which does not live up to its title is "The computer in archaeology" (Cloak, 1967). It gives almost no information about computer applications in archaeology. Cloak is an ethnographer; he indicates that his brief coverage is more or less complete up to 1967, which is completely incorrect (only a very small percentage of papers, mostly before 1963, have been mentioned). The paper has been extensively reviewed (Wilcock, 1969; Cowgill, 1970), with similar remarks. What

little is covered concerns the application of seriation to recent ethnographic assemblages, and there are some suggestions of approximation techniques which could save computer time.

1.4.16

Craytor and Johnson (1968) propose some refinements in computerised item seriation. Their software incorporates a new definition of seriation and a new criterion for ranking success in seriation for the various row and column permutations of a given matrix. The normal Robinson coefficient is formed by summing the coefficient differences from the outside edges of the matrix towards the diagonal, and is a maximum for the best ordering. Craytor sums the coefficient differences, then divides by the number of coefficients $n(n-1)/2$ and by the standard deviation of all the similarity coefficients. This value tends to the value 2 as the best ordering is approached.

1.4.17

Tugby (1969) provides a useful summary of statistical methods in archaeology. He describes matrix ordering methods, and defines 6 techniques:

Incidence matrix

O: site - attribute
P: attribute - site
Q: artefact - attribute
R: attribute - artefact
S: artefact - site
T: site - artefact

Square matrix

site - site (on attributes)
attribute - attribute (on sites)
artefact - artefact (on attributes)
attribute - attribute (on artefacts)
artefact - artefact (on sites)
site - site (on artefacts)

The Q and R methods are most useful in archaeology.

1.4.18

Sibson (1971) presents some thoughts on sequencing methods. He makes the point that all sequences must be regarded as reversible. Considering the formulation of dissimilarity coefficients, he criticises the different methods of calculating dissimilarity coefficients

used by Hole and Shaw (1967, see paragraph 1.4.14) since they seem to imply that the ordering will be changed if the different mathematical methods are employed. Sibson proposes that the seriation method should take account of the rank order of the dissimilarity coefficients rather than their relative magnitudes, a similar criterion to that employed in multidimensional scaling. The *discordance measure* of Hole and Shaw adds 1 to the error count if $j < i$ and $d(a_i, a_j) < d(a_i, a_{j+1})$ or if $j \geq i$ and $d(a_i, a_j) > d(a_i, a_{j+1})$, where i is the row number, j the column number, and $d(a_i, a_j)$ is the dissimilarity coefficient between items a_i and a_j , placed in row i , column j . This method, however, does not take into account the relative importance of errors in sequence in different areas of the matrix. To remedy this situation Sibson proposes an order-based measure which is sensitive to ordering errors, and which attaches more importance to sequence errors away from the diagonal than to those near the diagonal. The multidimensional scaling (Horseshoe) method of Kendall (q.v. paragraph 1.6.7) has been used to indicate sequences, but it is really designed to give an overall picture of the layout of a system through time, and does not yield a sequence except by an artifice (reading radially around the Horseshoe). Sibson shows that a relocation procedure based on his new discordance measure improves on Horseshoe output in some cases. The method is a valuable complement to the Horseshoe method.

1.4.19

Seriation is the purpose of FORTRAN V software for the UNIVAC 1108 machine developed by de la Vega (1972). The program constructs a classification of objects and seriates the classes using the Brainerd-Robinson technique. Each class in the first seriation is then seriated, so that the final seriation consists of single objects.

1.5 *Generation of Hypotheses and Models*

The use of models in archaeology has become very popular, but it was not until the advent of the computer, and in particular 3rd generation computers with large storage and fast arithmetic procedures, that the method became practical. In essence, the method consists of entering large amounts of archaeological data into the computer and employing "artificial intelligence" type algorithms to formulate hypotheses about the data or test models of a cultural entity. The method has been employed in particular for the discovery of artefact types from large bodies of data (using numerical taxonomy) and on large area surveys.

1.5.1

Spaulding (1953) describes the use of the fourfold (2 x 2) table, which has already received mention in paragraph 1.3.1.4.2. The *Kroeber coefficient of association* for such a table is

$$\frac{(a + d) - (b + c)}{a + b + c + d}$$

This coefficient is a measure of the association of the attribute represented by the left-hand column (elements a,c) with the attribute represented by the top row (elements a,b). Rows might represent, for example, surface finish of pottery, and columns the body materials. The coefficient may vary from +1.0 to -1.0 and is a measure of whether one or two types of pottery are to be recognised. The extreme values indicate two types of pottery, and a zero value indicates a single type only. No account is taken of sampling errors. For a total sample of n,

$$\chi^2 = \frac{n (ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}$$

There is one degree of freedom in a fourfold table. If x,y,z are the actual numbers of artefacts found with particular features

out of a total population of n , the expected occurrence of the triple association of these three features would be

$$e = n \left(\frac{x}{n} \right) \left(\frac{y}{n} \right) \left(\frac{z}{n} \right) = \frac{xyz}{n^2}$$

If the actual occurrence is 1, then the deviation is

$$D = 1 - e$$

The probability of the expected occurrence e is

$$p = \frac{xyz}{n^3}$$

Then

$$\chi^2 = \frac{D^2}{np(1-p)} \quad \text{with one degree of freedom.}$$

This analysis indicates a method of discovery of pottery types and the testing of their significance. All possible combinations of attributes must be tested in turn to detect all possible pottery types.

1.5.2

Gardin (1958) describes the use of punched card files in the descriptions of artefacts. As an example, metal tools are classified under the following headings:

- a) general (functional type, hafting, dimensions)
- b) functional part (sections, profiles)
- c) hafting part (sections, profile, distinctive details)
- d) connections between functional part and hafting part
- e) miscellaneous (ridges, other details, decoration).

These files were intended by Gardin to be the basis for large-scale analysis schemes, but it is now recognised that the computer is necessary for the scale of availability of information envisaged by Gardin, not just decks of punched cards.

1.5.3

Spaulding (1960) describes the use of the χ^2 test in the

statistical description and comparison of artefact assemblages.

1.5.4

Binford, L.R. (1964) discusses sampling methods. In the recording of archaeological information on files it may be necessary to record only a representative subset of the information because of the large bulk of repetitive data available. Under these circumstances the choice of sampling methods used to select the subset of data becomes important.

1.5.5

Work on the use of statistical methods in the analysis of large archaeological collections has been carried out in Russia (*Kovaļevskaja*, 1965).

1.5.6

With the availability of second generation computers the number of published papers begins to increase. Chenhall (1966) discusses the logic of models used for processing archaeological data on computers.

1.5.7

Martin (1966) reports the initial stages of IBM 7094 software to document cultural processes and to formulate laws of cultural regularity. The programs are for the investigation of the evolution of social organisation in prehistoric Arizona.

1.5.8

Millon (1966), in association with Cowgill, reports FORTRAN II and IV software for the IBM 7094 machine, and the SYMAP software for mapping. The programs are for the interpretation of the urbanisation at Teotihuacan, Mexico, on the basis of detailed mapping, surface reconnaissance, and selected excavations. Statistical techniques are used to order and interpret the data.

1.5.9

Rao (1966) uses the methods of estimating a discriminant function, originally introduced by Fisher, to decide between two simple hypotheses on the basis of observed data. The paper extends the idea to handle composite hypotheses. Rao, however, does not *generate* hypotheses, since Discriminant Analysis begins with a hypothesis.

1.5.10

The first of many papers by Doran (1967), a one-time member of the Department of Machine Intelligence and Perception, Edinburgh, puts forward a computer scientist's viewpoint on the use of computers in archaeology. He remarks that in the foreseeable future the computer-archaeologist should obtain more flexible and informal computing facilities, while the conventional archaeologist who only occasionally has a use for a computer will find more computer scientists to guide him, and may even find that he can use the machine himself *while it guides him*. This paper is a lucid presentation in laymen's language of the use of a computer to perform tedious calculations, clustering, multidimensional scaling, factor analysis, information storage and retrieval.

1.5.11

Cowgill and Millon (Cowgill, 1968; Millon and Cowgill, 1968) continue their report of the Teotihuacan project. The software is written in FORTRAN II and IV for the IBM 7094 and IBM 1620 machines. The geographic distribution of surface finds from 5,000 separate areas is to be analysed. Counts of many categories of artefacts and observations of attributes are punched on cards and transferred to magnetic tape. Statistical methods in use are fourfold tables for cross-tabulation of pairs of variables, or of variables versus locality, and multidimensional scaling. Localities which appear similar are considered to have been used by the same kind of people

for similar purposes. Graphics may be used to provide a visual output. Different surface collecting techniques were tested and found to have only minor effects on the results. Collection areas are based on the concepts of tract, site, complex, macrocomplex and neighbourhood and are not uniform in size. The various units within an area are linked by a ring structure, and recorded in the coordinates of the central mapping scheme. Some doubts are expressed about the use of the computer for all routine site recording. The software has facilities for recording, deletion, simplification and amalgamation of variables. The routine SYMAP is used for the production of distribution maps on the line printer. This FORTRAN IV routine was developed at the Laboratory for Computer Graphics, Harvard, and allows the mapping of areas of any shape, using interpolation between the specified data.

1.5.12

Several papers of interest to this topic appear in the collective volume edited by Binford and Binford (1968).

1.5.12.1

Hill (1968) examines the Broken K Pueblo site, in East-Central Arizona. Random samples are taken from the 95 rooms of the site, and the finds analysed by statistical techniques. χ^2 , the Fisher exact test and factor analysis are performed on the IBM 7094. The following attributes are used to compare rooms:

- a) size
- b) presence or absence of firepits
- c) presence or absence of mealing bins
- d) presence or absence of ventilators
- e) presence or absence of doorways
- f) height of door sill
- g) masonry style.

Ceremonial rooms (kivas) are removed from the sample because of their small number and special purpose. For the remainder, a bimodal room size is found ("large" or "small"). It was found that only one small room had a firepit. Performing a χ^2 significance test, there is less than .001 probability that the absence of firepits in small rooms is not significant. Similarly, there is less than .05 probability that the presence of mealing bins in large rooms is not significant, and less than .02 probability that the presence of ventilators in large rooms is not significant. Thus the rooms may be classified as kivas; large rooms with firepits, mealing bins and ventilators; and small rooms, perhaps for storage. Some types of pottery were for ceremonial use only. The Broken K Pueblo site shows good agreement with the Carter Ranch Pueblo (see paragraphs 1.3.8 and 1.5.12.2).

1.5.12.2

Longacre (1968) describes the analysis of the Carter Ranch Pueblo site, East-Central Arizona, by computer:

- a) UNIVAC (Brown and Freeman, 1964, see paragraph 1.3.8), using χ^2
- b) IBM 7094, using multiple regression.

It is deduced that the different types of pottery defined by the computer were associated with different activities, e.g. ceremonial use, cooking. Distinct activity areas, clusters of rooms with similar pottery types, are isolated.

1.5.12.3

Williams (1968) gives the theory for testing heterogeneities for dichotomous variables. In the fourfold table:

relative frequencies		p	q
		A	A'
r	B	pr	qr
s	B'	ps	qs

the Boolean variable A and its complement A' occur with relative frequencies p and q, and the Boolean variable B and its complement B' with relative frequencies r and s.

The frequency of $A \oplus B = A'.B \vee A.B'$ is $ps + qr$

The frequency of A.B is pr and of A'.B' is qs

Also $p = 1-q$, $r = 1-s$, and $pr + qr + ps + qs = 1$

If all M populations of equal size are identical with regard to p, q, r, s then the total frequency of A.B is

$$\frac{1}{M} \sum_{i=1}^M p_i r_i = \bar{p} \bar{r}$$

where \bar{p} and \bar{r} are the mean values of p and r. Also the total frequencies of A'.B' and $A \oplus B$ are $\bar{q}\bar{s}$ and $\bar{p}\bar{s} + \bar{q}\bar{r}$ respectively. Now for A'.B'

$$\bar{q}\bar{s} = \left(\frac{\sum_{i=1}^M q_i}{M} \right) \left(\frac{\sum_{i=1}^M s_i}{M} \right)$$

only when $q_1 = q_2 = \dots = q_M$
and $s_1 = s_2 = \dots = s_M$

i.e. there are conditions of homogeneity. If either of these conditions does not hold, then

$$\frac{\sum_{i=1}^M q_i s_i}{M} < \frac{\sum_{i=1}^M q_i}{M} \frac{\sum_{i=1}^M s_i}{M}$$

The magnitude of this inequality may be defined in terms of the covariance

$$\begin{aligned} \sigma_{qs} &= \frac{1}{M} \sum_{i=1}^M (q_i - \bar{q})(s_i - \bar{s}) \\ &= \left(\frac{1}{M} \sum_{i=1}^M q_i s_i \right) - \bar{q}\bar{s} - \bar{q}\bar{s} + \bar{q}\bar{s} \\ &= \left(\frac{1}{M} \sum_{i=1}^M q_i s_i \right) - \bar{q}\bar{s} \end{aligned}$$

$$\therefore \frac{1}{M} \sum_{i=1}^M q_i s_i = \bar{q}\bar{s} + \sigma_{qs}$$

Now $\sigma_{qs} = 0$ when the sample is homogeneous.

$$\therefore \frac{1}{M} \sum_{i=1}^M q_i s_i = \bar{q}\bar{s} \text{ when the sample is homogeneous.}$$

Similarly for A.B

$$\frac{1}{M} \sum_{i=1}^M p_i r_i = \bar{p}\bar{r} + \sigma_{pr}$$

$$\begin{aligned} \text{Also } \sigma_{qs} &= \left(\frac{1}{M} \sum_{i=1}^M q_i s_i \right) - \bar{q}\bar{s} \\ &= \left(\frac{1}{M} \sum_{i=1}^M (1-p_i)(1-r_i) \right) - (1-\bar{p})(1-\bar{r}) \\ &= 1 - \bar{p} - \bar{r} + \left(\frac{1}{M} \sum_{i=1}^M p_i r_i \right) - 1 + \bar{p} + \bar{r} - \bar{p}\bar{r} \\ &= \left(\frac{1}{M} \sum_{i=1}^M p_i r_i \right) - \bar{p}\bar{r} = \sigma_{pr} \end{aligned}$$

To summarise, the expected frequencies for a non-homogeneous sample are

$$\bar{p}\bar{r} + \sigma_{pr} \quad \text{for A.B}$$

$$\bar{q}\bar{s} + \sigma_{pr} \quad \text{for A'.B'}$$

and deriving a similar expression for A \oplus B we obtain

$$\begin{aligned} &\bar{p}\bar{s} + \bar{q}\bar{r} + \sigma_{ps} + \sigma_{qr} \\ &= \bar{p}\bar{s} + \bar{q}\bar{r} + \left\{ \left(\frac{1}{M} \sum_{i=1}^M p_i s_i \right) - \bar{p}\bar{s} + \left(\frac{1}{M} \sum_{i=1}^M q_i r_i \right) - \bar{q}\bar{r} \right\} \\ &= \bar{p}\bar{s} + \bar{q}\bar{r} + \left\{ \left(\frac{1}{M} \sum_{i=1}^M p_i (1-r_i) \right) - \bar{p}(1-\bar{r}) + \left(\frac{1}{M} \sum_{i=1}^M (1-p_i)r_i \right) - (1-\bar{p})\bar{r} \right\} \\ &= \bar{p}\bar{s} + \bar{q}\bar{r} + \left\{ \bar{p} - \left(\frac{1}{M} \sum_{i=1}^M p_i r_i \right) - \bar{p} + \bar{p}\bar{r} + \bar{r} - \left(\frac{1}{M} \sum_{i=1}^M p_i r_i \right) - \bar{r} + \bar{p}\bar{r} \right\} \\ &= \bar{p}\bar{s} + \bar{q}\bar{r} - 2 \left\{ \left(\frac{1}{M} \sum_{i=1}^M p_i r_i \right) - \bar{p}\bar{r} \right\} = \bar{p}\bar{s} + \bar{q}\bar{r} - 2 \sigma_{pr} \end{aligned}$$

1.5.13

Rowlett and Robbins (1969) report FORTRAN software for the IBM 360 machine for the selection of representative sites in archaeology. The purpose of the software is to determine which cemeteries and habitation areas of the Marne culture in Champagne, France are most typically representative of the culture during the La Tène Ia phase, and also which sites are the most representative of the four constituent Sonder groups of the culture. The contents and spatial relationships of the deposits are submitted to factor analysis.

1.5.14

Jardine (1970) points out the confusion which has arisen between *algorithms* and the *methods* which they implement, with particular reference to the Cambridge/Australia controversy reported in paragraph 1.3.12. The term "*model*" is also used in two quite different ways. Some authors (e.g. Lance and Williams) have distinguished two main types of algorithm, *divisive* and *agglomerative*. But the *single-link* clustering method, the only method satisfying the Cambridge criteria of adequacy, may be implemented by a divisive algorithm, an agglomerative algorithm, or by an algorithm which belongs to neither category. The term "*model*" has been used to cover the mathematical frameworks within which it is possible to analyse the methods. The other use of "*model*" has been to describe algorithms in terms of their application to some interpretation of the data (*analogue models*). Common analogue models have been: a distribution of points in n-dimensional Euclidian space; and a graph where the items are nodes, and values of the dissimilarity coefficients which are less than or equal to some threshold value are drawn in as edges. Geometric models have also been used for metric data. A rigorous observation of the distinction between methods and algorithms, and between mathematical models and analogue models

would seem to be desirable in the field of data simplification.

1.5.15

A series of papers by Doran (Doran, 1970, two papers; Doran 1971; Doran and Hodson, 1971; Doran, 1972) describe the development of general machine intelligence concepts applied to archaeology.

1.5.15.1

An exploration of systems theory and computer simulations applied to archaeology (1970) describes and evaluates recent attempts by archaeologists (e.g. Clarke, in his book *Analytical Archaeology*, 1968) to make use of concepts drawn from systems theory and cybernetics. It is suggested by Doran that these subjects may not have much to offer the archaeologist, and that the use of computer programs to simulate dynamic systems may be more promising. A suggestion of this thesis is that the *state tables* and *state diagrams* used in the design of sequential logic circuits for computers may be useful in describing archaeological systems whose evaluation depends not only on the environment (present inputs) but on the previous history (past inputs) of the system. The definition of a sequential circuit is a circuit whose current outputs depend not only on the present inputs but on the previous history (past inputs) of the circuit.

1.5.15.2

A comparison of archaeological reasoning and machine reasoning (1970) discusses the application of cluster analysis and multidimensional scaling. Doran argues that the process is a gross approximation to the true situation, because it is limited by the subjective assessments of similarity made by the archaeologist in the first place. A specific excavation report (Rainsborough Camp, near Banbury, Oxfordshire) is examined to establish the role of cluster analysis and other forms of reasoning, especially in relation to the determination of the

chronology of the site. A method is suggested (the method of *maximum likelihood*, see Hoel, 1962, and paragraph 1.4.5) whereby the process may be formalised and computerised. This method is related to non-numerical heuristic (learning) techniques.

1.5.15.3

Doran (1971) describes the computer analysis of a La Tène linear cemetery at Münsingen-Rain. The linear form of the cemetery suggests that it may have been built up starting at one extremity and continuing to the other, giving an approximately linear relationship between the date of a burial and the geographical location of the grave in the cemetery. Data concerns location, orientation and form of grave; sex and age of burial; location of skeleton and grave goods within the grave; details of grave goods (there are many fibulae, or brooches); and miscellaneous facts e.g. disease, planking. Hodson (1969, see paragraph 1.3.23) has carried out a conventional study. The analysis of 350 bronze fibulae has been carried out by computer. The computer study is in two parts, the calculation of similarity coefficients and the classification algorithm. Doran reviews the available methods: seriation, using abundance and binary incidence matrices; hierarchical systems; heuristic methods and hypothesis generation techniques, with interaction between computer and archaeologist. The fibulae data consists of measurements and motifs, and the results are presented in distribution map and histogram form.

1.5.15.4

Doran and Hodson (1971) describe "parallel" programs to generate and evaluate complex hypotheses. These programs must be designed to recognise the regular features of specific types. A first very simple program is working along these lines. They express the opinion that the techniques likely to be of the most use to archaeologists are only

just beginning to be developed. The techniques will be non-numerical as often as they are numerical, and they will be "tools" in organised "tool-kits" designed for use with computer-based data banks (a similar view is expressed by Wilcock, 1971, in the proceedings of the Mamaia conference, see paragraph 6.2.11). These data banks will sometimes be derived from individual excavations, and at other times will be much more widely derived. This article on the use of the "new statistics" in archaeology has received some criticism by Benfer, on the grounds that some complex statistical techniques have been used as a "cure-all", when a simpler technique would have sufficed. Doran (1972) replies to the criticism by agreeing that principal component analysis or Gower's principal coordinate analysis might be more appropriate than the indiscriminate use of multidimensional scaling for some applications, particularly when computing facilities are limited. However, k-means cluster analysis, which is a form of automatic hypothesis generation, and non-numerical techniques generally might prove more fruitful. The special demands of archaeological problems are the key role of classification procedures and the great diversity of non-numerical evidence which can be related to a problem.

1.5.16

Ammerman (1972) in collaboration with Cavalli-Sforza has investigated the diffusion of farming in the Neolithic and has detected what seems to be a regular rate of advance. He postulates a wave-front advance model, the mathematics of which have been refined by Kendall. Computer simulations have shown that given the conditions required by the model, viz. population growth and local migratory activity, a wave of advance sets in. The model makes the prediction that the expansion will take place at a constant rate, and in practice the rate of advance of the farming population is found to be approximately 0.6 miles per year. A wavefront advance model for a new gene was presented by Fisher in 1937 (Annals of Eugenics).

1.6 *Multidimensional Scaling*

The *multidimensional scaling* method arranges points representing archaeological units in n-dimensional space so that similar units are positioned close together and dissimilar units are positioned far apart. The method is non-metric, i.e. it takes no account of the absolute values of the dissimilarity coefficients, but only considers the relative values, sorting the coefficients into rank order of dissimilarity. The points are then positioned so that the distances apart are also in a rank order which corresponds, as far as possible, to the rank order of dissimilarities. The goodness-of-fit of the point distribution to the actual dissimilarity coefficients is calculated and expressed as the *stress*. The best number of dimensions for the distribution (often 2 or 3 for archaeological applications) is also found. The points are repositioned for every iteration of the program, each point acting under a "*resultant*" force generated by its distance discrepancies relative to all the other points and the directions of the other points. The iterations and changes of dimensions continue until a satisfactory solution (small stress value) is obtained, or the stress value does not reduce quickly enough, or a specified number of iterations has been performed.

1.6.1

The input for the multidimensional scaling method is a half-matrix less diagonal of similarity or dissimilarity coefficients comparing each item with every other item in the distribution. The first archaeologist to make steps towards the construction of such a matrix was W.M. Flinders Petrie (1899). The paper describes Petrie's original work on Egyptian prehistoric remains, and Kendall has developed the method into the *Kendall-Petrie Concentration Principle*. The method was termed *sequence dating* by Petrie, but is now more-often called *Seriation* (see section 1.4). Petrie examined 900 pre-dynastic Egyptian graves containing about 800 types of pottery in all. The *incidence matrix*, which represents the occurrence of each type of pottery in each grave, contains information

concerning the sequential ordering of the graves in time and the time periods during which the different varieties of pottery flourished. The incidence matrix could be analysed by the Brainerd-Robinson method, but Kendall generates from it a square matrix which compares each grave with every other grave. Kendall postulates that this second matrix contains all necessary information for the ordering of the graves in n -dimensional space by the multidimensional scaling method. If the information is linear, as it is in the case of Hodson's Munsingen-Rain study, the points appear in a roughly horseshoe-shaped distribution in 2 dimensions, and the sequence of the graves can be obtained by the artifice of reading radially around the horseshoe.

1.6.2

The first computer program for the multidimensional scaling method was designed by Shepard (1962). The aim of the program is to determine a minimum set of cartesian coordinates for points related by some unknown, fixed monotonic function of the distance between them. The distances between the points in the initial configuration are calculated, placed in rank order (smallest distance has first rank), and compared with the ranking already obtained from the proximity measures input. If there are n points in all, $n-1$ vectors are formed from each point to the other $n-1$ points, either pointing towards the other points or away from them depending upon whether the rank of the distance between the two points is too large or too small. The length of each vector is determined by its discrepancy in rank. The $n-1$ vectors are interpreted as $n-1$ forces which are tending to pull a point towards those points that are too distant and away from those which are too close. The n points are then "simultaneously" displaced according to the vector sum (resultant) of their $n-1$ vectors. For the initial configuration, coordinates can always be found in $n-1$ dimensions such that all $n(n-1)/2$ interpoint distances are equal, thus introducing no bias

into the final solution (e.g. 3 points may be placed at the vertices of an equilateral triangle in 2 dimensions, 4 points at the vertices of a tetrahedron in 3 dimensions, etc.). The dimensional space is collapsed by the algorithm to the minimum number of dimensions giving an acceptable stress in the final configuration.

1.6.3

Kendall (1963) gives his first exposition of Petrie's sequence dating, and the statistical problem raised by the method. He treats the problem entirely as one of estimation. A scoring function is formulated which is a measure of discordance in the order of the graves, which must be minimised over all possible permutations. For any reasonable number of graves the number of permutations becomes prohibitive, and for this reason a two-stage reduction procedure using similarity functions is proposed. The scoring function used

$$S(\Pi) = \sum_i N_i \log R_i$$

where Π is the permutation

N_i is the total number of the i^{th} variety of find

R_i is the range of the i^{th} variety of find (in number of consecutive graves in the present order)

1.6.4

The main colleague of Shepard in the study of multidimensional scaling is Kruskal. Kruskal gives the theory behind Shepard's practical implementation in two 1964 papers, and also describes another computer program (MDSCAL) which has received world-wide application. Multidimensional scaling is the solution of the problem of representing n objects geometrically by n points so that the interpoint distances correspond in some sense to the experimental dissimilarities between objects. Kruskal's fundamental hypothesis is that the

dissimilarities and distances are monotonically related. A quantitative, intuitively-satisfying measure of goodness-of-fit to a non-metric hypothesis is defined. Stress and goodness-of-fit are classified as follows: 20% poor; 10% fair; 5% good; 2½% excellent and 0% "perfect" (meaning that there is a perfect monotonic relationship between the dissimilarities and distances). The progress of the algorithm may be shown by scatter diagrams in which distance is plotted against dissimilarity for all $n(n-1)/2$ pairs of objects. Thus if the dissimilarities are placed in ascending order, the distance variable should always increase for a perfect monotonic relationship. In practice this is not usually attained, and points are fitted to an ascending curve defined by a monotonic series of numbers as nearly as possible equal to the actual distances. Deviations are measured along the distance axis, and the stress is defined as

$$\sqrt{\frac{\sum (d_{ij} - \hat{d}_{ij})^2}{\sum d_{ij}^2}}$$

where the d_{ij} the actual distances apart of the points and the \hat{d}_{ij} are the monotonic ascending values. The division by $\sum d_{ij}^2$ is to normalise the expression so that it shall be invariant under expansion. The configuration required is that with minimum stress and this is found by altering the distances, moving point i nearer to point j if $\hat{d}_{ij} < d_{ij}$ and further apart in the opposite case, starting from an arbitrary configuration. Typically between 15 and 100 iterations are necessary. Each point is subject to many such motions simultaneously, and usually these will be in partial conflict. The problem is to minimise a function of many variables, and the "*method of steepest descent*" is used. Examples of the process are given, with illustrations of output obtained from a microfilm printer/plotter. An arbitrary initial configuration used is to arrange the points at equal spacing along two lines at right

angles (bottom and left-hand side of the diagram). Scatter diagrams are given for the initial configuration and after 10 and 50 iterations. The choice of the number of dimensions is governed by the minimum number which produces acceptable stress and where there is little significant improvement when the number of dimensions is increased. For n points there will always be a perfect fit with $n-1$ dimensions. The problem of missing dissimilarities is handled by omitting the corresponding items from the analysis entirely. Recently function minimisation techniques have been developed as an alternative to the method of steepest descent.

1.6.5

Ellisseeff (1965) reports the application of multidimensional scaling to ancient Chinese bronzes.

1.6.6

Hodson, Sneath and Doran (1966) describe the application of multidimensional scaling, among other methods, to the linear cemetery at Munsingen-Rain. This linear relationship between time and geographical location in the cemetery causes the scalogram to exhibit a horse-shoe shaped distribution of points.

1.6.7

Much of the further development of the method has been carried out by Kendall and his associates. The work is reported in Kendall (1969); Kendall (1970), two references; Kendall (1971); Wilkinson (1971); and Hiorns (1971).

1.6.7.1

Kendall (1969) defines the *incidence matrix* expressing the varieties of item present in different graves, and the two square matrices which may be derived from it, one of them giving the number of graves in which both the i^{th} and j^{th} varieties are both represented (V matrix), and the other giving the number of varieties which are held in common by graves i and j (G matrix). It is emphasized that

graves represent points in time, but varieties may precede, succeed, overlap, lie wholly within or wholly without other varieties.

"*Petrifaction*" is the permutation of the rows and columns of a square matrix to bunch the high values of the coefficients together. The use of Kruskal's MDSCAL, modified at Cambridge to give plotter output and called HORSHU, is demonstrated on the G matrix. It is shown that the permuted form of the G matrix is in Robinson form. The *Kendall-Petrie Concentration Principle* is stated as "if the typology is chronologically significant, and when the graves have been correctly ordered (or anti-ordered), then the sequence date ranges for the individual types will be found to have been individually or in some communal way minimised". The use of MDSCAL on the V matrix would not be appropriate, since what is required in this case is a program to produce a model of overlapping segments on a line or overlapping discs on a plane. The use of MDSCAL on Hodson's Münsingen-Rain data is described.

1.6.7.2

Kendall largely reiterates his 1969 remarks in a 1970 paper. The horseshoe distribution may be used to indicate time sequence by reading radially around the horseshoe. MDSCAL is strictly not a serialisation method, however. A MDSCAL map of Romania is illustrated which was generated by supplying as data the *rank order* of road distances between towns (*not* the actual road distances). Hodson's Münsingen-Rain data are illustrated by a horseshoe plot with Hodson's numbering (purposely scrambled to give no indication of the true order to the mathematicians) and a true map of the cemetery with numbering predicted by MDSCAL. There is very good agreement between the computer numbering and the actual order in the cemetery. It would be useful to "unbend the horseshoe" in order to give a linear representation with time, and this may be done by reducing the weight given to highly dissimilar pairs (Needham, 1965). Wilkinson realised that the travelling salesman problem also has a bearing on this, and success in the method is reported in

Kendall's note in *Science and Archaeology* (1970). The method of unbending the horseshoe is so successful that it seems that the Münsingen-Rain data is truly one-dimensional.

1.6.7.3

Kendall in his 1971 paper uses the term *abundance matrix* in which element ij contains a non-negative number specifying the frequency of occurrence of the j^{th} variety in the i^{th} grave. This frequency may be the actual number of occurrences of variety j in grave i or the proportion of the total contents of grave i represented by variety j . The abundance matrix yields the G square matrix which may be submitted to multidimensional scaling. The "horseshoe" graphical output (which is only a horseshoe if a linear relationship is inherent in the data) may be given added meaning by linking by vectors all pairs of points whose similarity coefficients exceed a critical value. A similar procedure (but using Renfrew-Sterud linkage) has been employed in this thesis (see paragraph 1.10.1, Figures 1.2, 1.4, 1.6). The essential linear data from the Münsingen-Rain cemetery produces a rough horseshoe with the points ordered in about the same order predicted by Hodson's conventional study, and also roughly in the geographical order of the linear cemetery. Wilkinson's "*circle up*" method of unbending the horseshoe is briefly described. The matrix is "circled-up" before the application of MDSCAL and then produces a more linear plot. The routine DIMDROP telescopes the dimensions of the result by projecting the results onto the principal component. There are storage problems in handling more than 50 graves on most computers as well as penalties in the time taken for the run. During the iterations of MDSCAL it is also possible that points will collide under the action of their resultant vectors. The original MDSCAL program terminates under such circumstances, but the Cambridge version of MDSCAL, called HORSHU allows points to "pass through" one another during the iterations.

Other modifications to the standard MDSCAL program are given in an appendix to the paper. For input of a binary incidence matrix which is *sparse*, i.e. which consists mainly of absences of varieties (0) and which has very few presences (1), a coding method with automatic check is presented. For example a line of a six-column matrix containing 1s in column 1 and 3 only is coded as 1 3 7 where the 7 is an end-of-line check.

1.6.7.4

Wilkinson (1971) was the first to recognise the applicability of *minimal Hamiltonian circuits* (i.e. the *travelling salesman problem* of operations research) to the extraction of a chronologically ordered sequence from the output of MDSCAL, which is strictly not a seriation method. He starts off by standardising the matrix, by removing rows of all zeros and duplicate rows. This is a similar procedure to that used at the beginning of the *Roth technique*, a method of generating efficient sets of tests in diagnostic programming. The remaining m rows and n columns of the matrix are considered to be m points in n -dimensional space. The distance between points r and s is calculated as

$$\begin{aligned} & (\text{number of 1s in } r^{\text{th}} \text{ row not in corresponding positions of the } s^{\text{th}} \text{ row}) \\ & + (\text{number of 1s in } s^{\text{th}} \text{ row not in corresponding positions of the } r^{\text{th}} \text{ row}) \end{aligned}$$

i.e. the result of a "not equivalent" logical operation on the two rows followed by a bits count. This method of distance measure was used by Wilcock in 1967 in a seriation program written for the KDF9 computer, the order code for which conveniently contains the instructions NEV (not-equivalent) and BITS (bit count) (unpublished report, Computer Centre, University of Keele, 1968). A *Hamiltonian circuit* is a re-entrant path passing once only through the m points of minimum length $2n$. The travelling salesman problem is to find the shortest Hamiltonian circuit. Wilkinson also defines his "circle-up" operation which uses

the method to "unbend the horseshoe".

1.6.7.5

Kendall has applied MDSCAL to the data produced in the study of the Otmoor parish registers by Hiorns et. al. (1971). The method uses the number of marriages between persons from different pairs of parishes as a proximity measure. Before the advent of modern transport people would be expected to choose marriage partners mostly from their own parish or neighbouring parishes. The data is taken from the parish registers and forms the basis for two studies in marriage migration behaviour, c.1600-1850 and 1851-1967, the second period covering the time when the advent of the railway line to Oxford much increased the mobility of the Otmoor population. MDSCAL operating on the proximity measures has produced a surprisingly accurate map of the Otmoor region. Hiorns et.al. had no interest in drawing a map, since plenty of maps exist, but were concerned with the genetic consequences of marriage migration behaviour. Kendall is continuing to investigate the possible use of MDSCAL in map generation, applying the method to the layout of medieval settlements using agricultural records and parish registers. It may even be possible to locate "lost" medieval villages which disappeared at the time of the enclosures.

1.6.8

Kruskal (1971) emphasizes that multidimensional scaling is essentially not a seriation method, and that although time is expected to be one of the dimensions, there are other dimensions to consider. The non-metric hypothesis of MDSCAL is that there is error in the measurements of distance between objects, such that

$$f(d_{ij}) = d_{ij} + e_{ij}$$

where f is some order-preserving function such as

$$f(d) = e^d, f(d) = d^3 \text{ or } f(d) = \sqrt{d}.$$

It is possible to reconstruct the configuration approximately even though the function is not known. The number of dimensions is chosen to be the smallest number giving an acceptable stress. An "elbow" is often visible in the plot of minimum stress σ_v number of dimensions

curve, and the number of dimensions at the elbow is taken from the final configuration. This is the number of dimensions for which no significant improvement in the stress is achieved by increasing the number of dimensions further.

1.6.9

Chouraqi (1972) reports FORTRAN IV software for IBM 360 and UNIVAC 1108 machines to classify objects and generate scalograms. The similarity function F is defined in terms of the individual similarity coefficients $S(i,j)$ as

$$F = + 1 \text{ if } S(1,k) - S(1,j) < 0$$

$$F = 0 \quad \text{if } S(1,k) - S(1,j) = 0$$

$$F = - 1 \text{ if } S(1,k) - S(1,j) > 0$$

where $j \geq k$ and the lower half matrix only is considered. The Robinson method carried out on the matrix results in F taking up a maximum value, corresponding to the increase of similarity coefficients towards the diagonal. A second algorithm written in FORTRAN V generates a scalogram by ordering the objects according to the "centre of gravity" of the similarity coefficients of the corresponding row or column of the matrix. A "moment of inertia" is calculated for each "centre of gravity" and used to define a figure of merit. The ordering continues until a maximum of this figure of merit is attained, when the centres of gravity define the points of the scalogram.

1.7 Cumulative Percentage Graphs

The *cumulative percentage graph* has been used by some archaeologists to compare assemblages of artefacts. The types of artefact are arranged in the same order in all assemblages and the population of each type is expressed as a percentage of the total population for each assemblage. A graph (or more strictly, a histogram) is plotted for each assemblage, plotting the actual percentage for the first type, the sum of the first and second types for the second type, the sum of the first three types for the third, and so on, until 100% is attained with the last type. The dissimilarity of two assemblages is judged from the amount of discrepancy between the corresponding histograms. This method is inherently unreliable, however, since it has been shown that merely by altering the order of the types quite different subjective judgments of the dissimilarity between two assemblages may be obtained. A more reliable procedure would be to compare the actual (or cumulative) percentages by computer in a purely objective algorithm.

1.7.1

Statistical tables for the estimation of goodness-of-fit of two sample cumulatives have been published by Smirnov (1948).

1.7.2

Massey (1952) develops the theory and publishes tables of use in the interpretation of cumulative percentage graphs. The tables are of the function

$$d = \max | S_{N_1}(x) - S_{N_2}(x) |$$

where $S_{N_1}(x)$ is the proportion of the N_1 observed values x_i less than or equal to x , and $S_{N_2}(x)$ is a similar proportion for the second cumulative. Tables are given for N_1 and N_2 less than or equal to 10

and all possible values d_α and α , such that the probability that d is less than or equal to d_α is α .

1.7.3

Bohmers (1963) uses a graphical analysis of the percentage distributions of attributes in flint artefacts, although he plots the percentages as actual values, not cumulatives. Bohmers calls these *correspondence diagrams*.

1.7.4

Bordes (1964) criticises the correspondence diagrams of Bohmers, derogatorily calling them "crossword puzzles". Bordes prefers cumulative percentage graphs, but his arguments seem less than convincing.

1.7.5

Doran and Hodson (1966) use the cumulative percentage graph in conjunction with multidimensional scaling to analyse Palaeolithic flint assemblages. The least strain result is collapsed to one dimension to give a linear order for the assemblages.

1.7.6

Kerrich and Clarke (1967) criticise the work of Bordes and point out some dangers in the use of cumulative percentage graphs. Quite different graphs may be constructed from the same set of data purely by rearranging the order of the types of artefact. Possible errors which arise are sampling errors, the taking of percentages (which often hide significant aspects of the data and are no substitute for the actual data), and most important of all, perception errors for different observers.

1.7.7

Whallon (1968) uses the cumulative percentage graph to study late prehistoric social organisation in New York State. The maximum degree of variability produces a straight, diagonal line from 0% to 100% at n variables. Maximum homogeneity produces an almost vertical

line from 0% to 100% at the *first* variable, and a horizontal line from the first to the n^{th} variable at 100%. The subjective difficulties in judging similarities between cumulative percentage graphs is removed by calculating areas under the graphs by the trapezoidal method. The area under a homogeneous distribution

$$= \left(\frac{1}{2} \times 1 \times 100\right) + 100(n-1) \\ = 100\left(n - \frac{1}{2}\right)$$

and the area under the maximum variability curve

$$= \frac{1}{2} \times n \times 100 \\ = 50n$$

The ratio $v = \frac{\text{maximum variability area}}{\text{actual area}}$ ranges from 1.0 for maximum variability down to just greater than 0.5 for homogeneity. For intermediate graphs $1.0 \geq v > 0.5$ where v is the coefficient of variability. Alternatively a coefficient of homogeneity may be defined as

$$h = \frac{\text{actual area}}{\text{maximum homogeneity area}}$$

where $1.0 \geq h > 0.5$. A more complex expression for a coefficient of homogeneity used by Whallon is

$$h = 1 - \frac{2(100n - \sum y)}{100(N-1)}$$

where the y coordinates are the heights of the n points on a cumulative percentage graph, and N is a fixed maximum number of variables (chosen as 20 in Whallon's study). This introduces a weighting factor which favours those distributions with a smaller number of variables (since these are inherently more homogeneous). If this is not done inconsistent results may be obtained when graphs with different numbers of variables have to be compared.

1.7.8

Bordes and de Sonneville Bordes (1970) use cumulative percentage graphs to differentiate between Palaeolithic assemblages. They criticise the conclusions obtained by Binford and Binford using factor

analysis on similar assemblages, preferring the interpretation that the different assemblages represent different traditions rather than Binford's conclusion that they represent different activities carried on at different places by essentially the same kind of people.

1.7.9

Thomas (1971) points out that cumulative percentage graphs are analogous to the R technique, expressing type distributions in terms of assemblages. He objects to the use of cumulative percentage graphs in assessing similarities between assemblages (Q technique) on the grounds that the data has no inherent order, i.e. types may be reordered at will to produce quite different curves. Discrepancies in important variables may produce undue dissimilarity throughout the rest of the curve. Thomas emphasizes the advantages of using the computer to produce diagrams for publication, a proposition which has been developed at length in this thesis.

1.8 *Trend Surface Analysis*

Geographers have begun to apply statistics to their studies, and one of the techniques frequently used is *trend surface analysis*. Quantitative observations are taken at sampling points throughout the area under study and used to define a pseudo-3D-surface. The properties of this surface are then investigated statistically. Linear, quadratic, cubic and higher polynomial trends may be detected in various directions. It is a common practice to subtract the primary trend from the original surface and to carry out secondary trend surface analysis on the *residuals*. This process has been applied to the study of the movement and interplay of cultures.

1.8.1

The basic theory of the method is given by Krumbein (1959). He describes the trend surface analysis of contour-type maps with irregular control-point spacing, using software for the IBM 650 machine. The program fits linear, quadratic, cubic, quartic and higher polynomial trends to the surface data, and calculates residuals. The general cubic surface y may be defined in terms of two variables u, v by the equation

$$\begin{aligned} y = & a + bu + cv \\ & + du^2 + euv + fv^2 \\ & + gu^3 + hu^2v + juv^2 + kv^3 \end{aligned}$$

where a is a constant, b and c are linear coefficients, d, e and f are quadratic coefficients, and g, h, j and k are cubic coefficients. u and v are quantifications of archaeological entities, e.g. the occurrence of different pottery types.

1.8.2

Chorley (1964) explores some of the possibilities of the method in geography.

1.8.3

Sneath (1967) applies trend surface analysis not to the analysis of field data but to the *method of transformation grids* used to explore the relationship between different skull types and the evolution of man (based on numerical taxonomy). Starting with diagrams of

the skulls, corresponding points on each diagram (e.g. tip of nose, chin) are marked, and their coordinates are recorded. The diagrams are then scaled and rotated to give the best possible fit between a skull type and the next in the evolutionary series; this gives a measure of size and shape difference. Using these standardised diagrams, the displacements between each point and the corresponding point on the next diagram are measured, then submitted to trend surface analysis. The displacements are analysed in terms of linear, quadratic, cubic and higher order trends. The differences based on the trends alone can now be estimated. The scaling and fitting process involves centring (translation), reduction to the same size, determination that the profiles of both skulls face the same way, rotation, and estimation of fit; algorithms are given for these processes. The differences in x and y coordinates of the series of points are made the basis of the trend surface analysis. Various trends are distinguished: x trend in x (expansion in x), y trend in x (shear), y trend in y (expansion in y), and x trend in y (shear). Linear, quadratic and cubic surfaces are fitted, and interpreted as different changes between the species. Starting with a rectangular grid superimposed on the skull of *Homo sapiens*, distorted grids are fitted to the skulls of *Homo erectus*, an *Australopithecus* species similar to *Australopithecus africanus*, and *Pan troglodytes* such that the grid intersections identify corresponding points on all the skulls.

1.8.4

Berry and Marble (1968) edit a collection of studies in statistical geography. The use of 3D arrays giving "characteristics" as a function of places and time are described. The characteristics may be settlement patterns, resources, vegetation, etc. The use of computer models for the development of settlements is described. The production of maps by computer receives attention, and a *global reference code*, based on a latitude and longitude, is defined, whereby any point on the

earth's surface may be identified. Cartograms are defined, which are distorted views of map areas, each area being related to the population, market size, etc. Different methods of geographic sampling are described. The analysis of spatial distributions by trend surface analysis is covered. *Isarithmic* (Greek: same number) maps, i.e. maps based on *regular* sampling points, each at the centre of a defined hexagonal area, are described. In conventional isarithmic mapping a point is not allowed to influence other points outside the area defined by joining its encircling neighbour sampling points, but this is not maintained in the study described. Instead points are allowed to affect wider areas by means of "moving averages". The effect is to damp down local noise and to give a regional trend with local residuals. There are many different types of trend surface analysis available, with different sampling methods. It is found that a regular grid spacing acts as a filter, attenuating certain frequencies, and the results thus depend on the spacing of the sampling grid which is chosen. *Isopleths* (Greek: same fulness) are contour-type lines fitted to observations of the same value throughout the field. Care must be taken in the identification of "noise", since noise may in some circumstances have the same magnitude as local residuals.

1.8.5

Chorley and Haggett (1968) describe trend surface methods in the volume edited by Berry and Marble (1968) covered in paragraph 1.8.4 above.

1.8.6

Harbaugh and Preston (1968), also in the volume edited by Berry and Marble (1968), describe the use of Fourier series in geology. The use of the *trapezoidal rule* for the numerical integration of empirical data multiplied by sine or cosine is described. The method is analogous to the use of a *power series* polynomial in trend surface analysis. The

power spectrum for each harmonic is found by squaring and adding the coefficients of the sine and cosine terms for the harmonic. The use of double Fourier series analysis for areal studies is described. Major trends should be subtracted first, the Fourier analysis carried out on the residuals, then the trend superimposed.

1.9 *Manual Methods*

The availability of such computer techniques as numerical taxonomy, seriation, factor analysis and multidimensional scaling has encouraged archaeologists to submit data as a routine to such methods. Although the advantages to be gained have been pointed out above, such methods are not "cure-alls", and it is worthwhile to note that there are manual methods which yield similar results, and which take only a few minutes to perform on small matrices. The Renfrew-Sterud method in particular has advantages, and the use of this method has been combined with multidimensional scaling in this thesis (see paragraph 1.10.1). It is thought that this is a novel approach, the success of which may be judged from the scalograms for Samian pottery, British Beaker pottery and Eastern European projectile points (Figures 1.2, 1.4, and 1.6).

1.9.1

Renfrew and Sterud (1969) present two simple manual methods for the seriation and linkage of items from archaeological assemblages by *close-proximity analysis*. The method requires a matrix of similarity coefficients, as do other seriation methods. When the archaeological material may be seriated well, the Renfrew-Sterud methods produce structures which compare well with computer seriations of the same data. When several branch chains exist (see in particular the British Beaker pottery scalogram, Figure 1.4), computer seriation is inaccurate, for the seriation method artificially forces the assemblage into a linear order, and in fact the Renfrew-Sterud manual methods (which of course could be computerised) reveal more about the structure of the data. Also for the computer seriation method the time necessary mounts as the square of the matrix size, while the manual methods increase only linearly with matrix size. The "*double-link*" method (not to be confused with Sibson's double-link) requires that the two highest similarity coefficients in every row of the matrix be identified, and each item is

linked provisionally to two others. Any loops which appear are broken by removing the weakest link in each loop. The result is a linear or branched structure relating the items of the assemblage. The *descending coefficient* method represents the linkages by Venn diagrams, and a succession of discrimination levels are arbitrarily chosen to reveal the structure of the data. A similar, but not identical method has been used in the BASIC program PHENON written for this thesis, and the successive Venn diagrams may be added to the scalogram with Renfrew-Sterud linkage (see Figure 1.19). Renfrew and Sterud review the work of Sackett, Hole and Shaw, Kuzara, Mead and Dixon in the light of these new methods. An analysis is also performed for some data from Cycladic cemeteries.

1.9.2

Gelfand, in two similar 1971 papers, describes two manual methods for the seriation of linear data. Both methods require a half matrix of similarity coefficients. The *first method* arranges the $n(n-1)/2$ coefficients in decreasing order. The $n-1$ largest similarity coefficients are selected subject to the following rules:

- (a) coefficients in which one of the items is already linked by two higher coefficients are ignored;
- (b) coefficients in which both items are singly linked are also ignored.

The above rules ensure that in the $n-1$ coefficients selected $n-2$ items occur exactly twice and 2 items appear exactly once, and all items communicate, i.e. $i-k_1, k_1-k_2, \dots, k_{n-2}-j$ to permit seriation. The final seriation is produced by starting with a singly-linked item occurring in the $n-1$ selected coefficients and continuing by the indicated linkages until all n items are built into the linear sequence. The first method does not use all the information in the matrix. The *second method* overcomes this by ordering the coefficients in each row of the matrix in order of decreasing coefficients, producing a linear

order for each row of the matrix. For example, if 1,4 is the largest coefficient in row 1, write down 1-4. If 1,6 is the next largest, and $4,6 > 1,6$ then write 1-4-6, but if $4,6 < 1,6$ write 6-1-4 etc. If equality occurs then compare the new item with the next rightmost and leftmost members of the existing list, and position the new item accordingly. This process is carried out for every row of the matrix, then the ranks of all items are averaged over all the rows to produce the final ranking. The success of both these methods depends very heavily on the assumption that there is a single dimension underlying the data. The second method seems to give more reliable results. Both methods can give a good starting point for multidimensional scaling.

1.10 *Statistics Programs*

Several statistics programs have been developed for this thesis, written (with the exception of MDSCAL) either in ALGOL 60 with ICL 4130 input/output conventions, or in BASIC for use with multi-access remote teletype terminals. The programs will be described individually, with particular attention to original features. Flowcharts will in general be given only to describe original features. Full listings of the programs are available in the printout file.

1.10.1

ARCJWO18 MDSCAL Disc (FORTRAN)

This is a disc implementation of the Kruskal MDSCAL program. The implementation was tested with a set of Otmoor data with known results supplied by Kendall and Hiorns. The program allows the input to be in various forms: either similarity or dissimilarity coefficients; full matrix or half matrix; diagonal present or absent. Control cards may be used to specify the maximum number of iterations to be performed, and also the maximum and minimum number of dimensions for the analysis. Points representing the items are repositioned on every iteration under the action of resultant vectors as described in section 1.6 above. The process continues until an acceptable value of stress is achieved, or the improvement is too slow, or the specified number of iterations have been performed. The output is a history of the computation showing the change in stress on each iteration, the final number of dimensions and the stress achieved, and positions of all points in the scalogram in arbitrary coordinates. The full rank order of the coefficients is also output, each with its best hypothetical distance between the pair of items concerned, and the actual distance between the items in the final

positions. The program has been used to perform multidimensional scaling on the following new data:

- a) similarity coefficients for common types of Samian ware. The coefficients were generated by an original algorithm in the program ARCJW019 (see paragraph 4.4.3). The final configuration of 13 types in 2 dimensions has stress 0.010. The plain scalogram is illustrated in Figure 1.1 and with Renfrew-Sterud linkage in Figure 1.2. It may be seen that the Renfrew-Sterud linkage enhances the scalogram portrayal of the relationship between the various types of pottery, and it is claimed that the conjunction of these two techniques is original. The scalogram isolates two main groups of pots, the small "dish-type" pots (Drag.27,33,31,18/31,18,36,35) and the large "bowl-type" pots (Drag.37,29,45,38). Drag.30 and 68 are discordant, the first being cylindrical, and the second having a tall convex profile. The evolutionary sequence 18,18/31,31 is clearly seen.
- b) similarity coefficients for British Beaker pottery. The coefficients were generated by program ARCJW019 as above, and are based on the types classified by Clarke. The final configuration of 19 types in two dimensions has stress 0.107. The plain scalogram is illustrated in Figure 1.3 and with Renfrew-Sterud linkage in Figure 1.4. It may be seen how the Renfrew-Sterud linkage emphasizes Clarke's intrusive types *All-over-cord* (AOC), *European* (E), *Western/Middle-Rhine* (W/MR), *Northern/Middle Rhine* (N/MR) and *Northern/North Rhine* (N/NR). Clarke's intrusive types *East Anglian* (E.Ang) and *Barbed wire* (BW) are placed by the method between N/NR and the *Southern* (S1) type which lies centrally on the scalogram. Clarke's intrusive *Northern* type

MULTIDIMENSIONAL SCALING PLOT
COMMON SAMIAN TYPES
(SIMILARITY COEFFICIENTS)

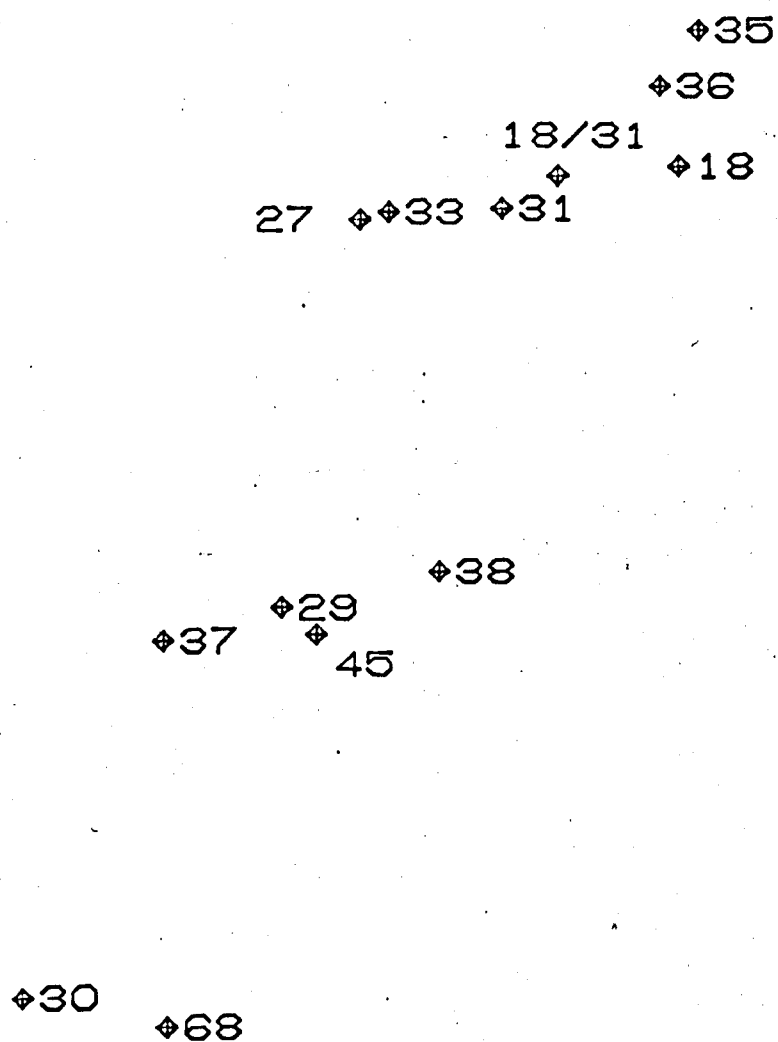


Figure 1.1

COMMON SAMIAN POTTERY TYPES
SCALOGRAM
WITH RENFREW-STERUD LINKAGE

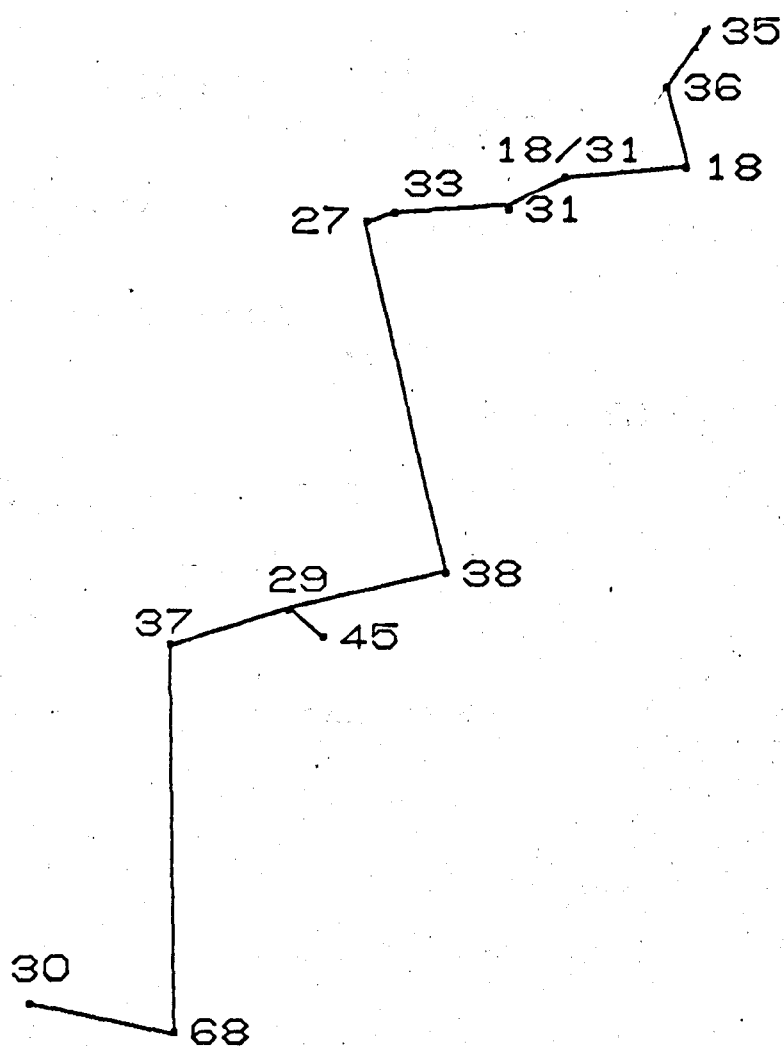


Figure 1.2

MULTIDIMENSIONAL SCALING

BRITISH BEAKER POTTERY

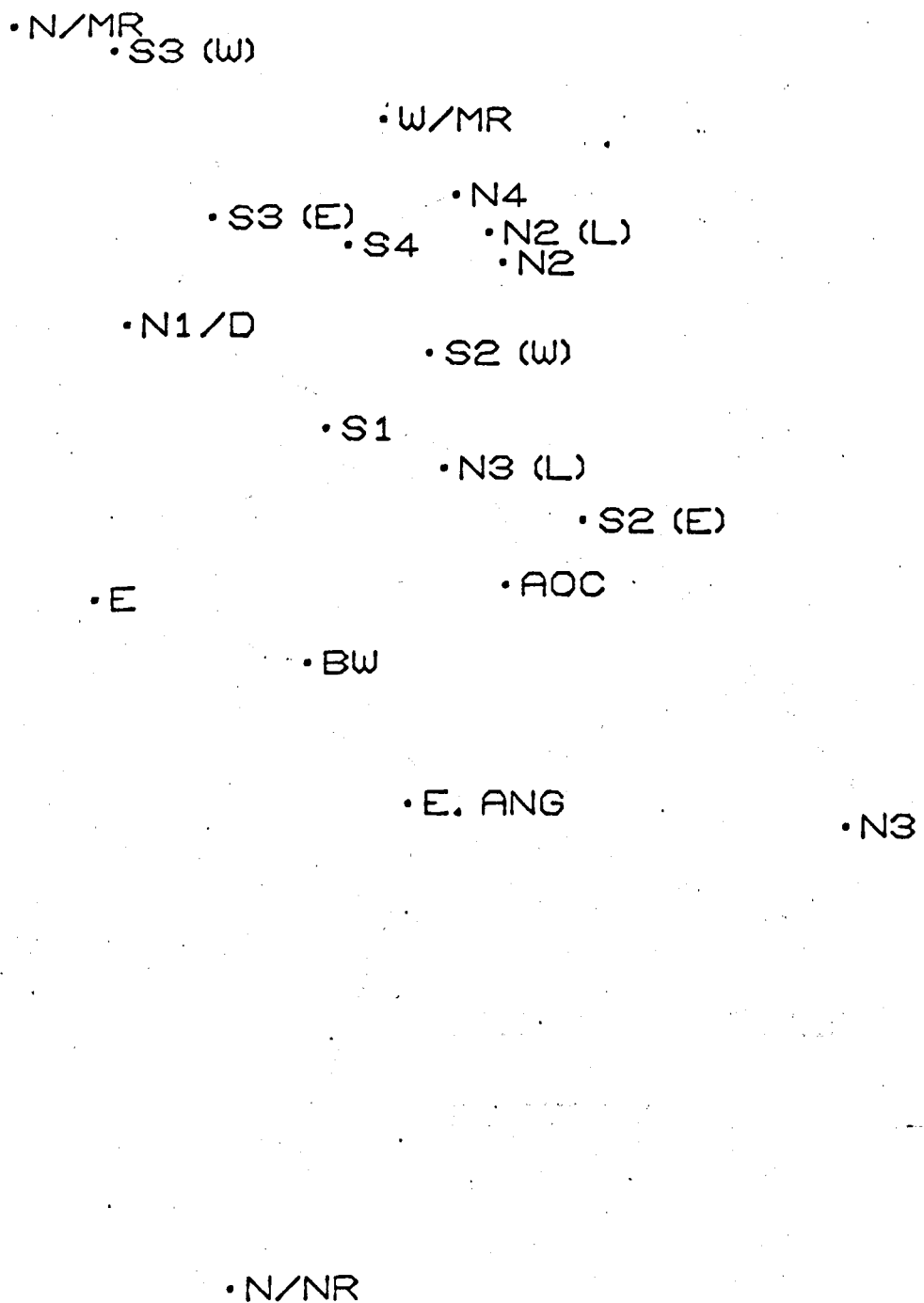


Figure 1.3

MULTIDIMENSIONAL SCALING WITH RENFREW-STERUD LINKAGE

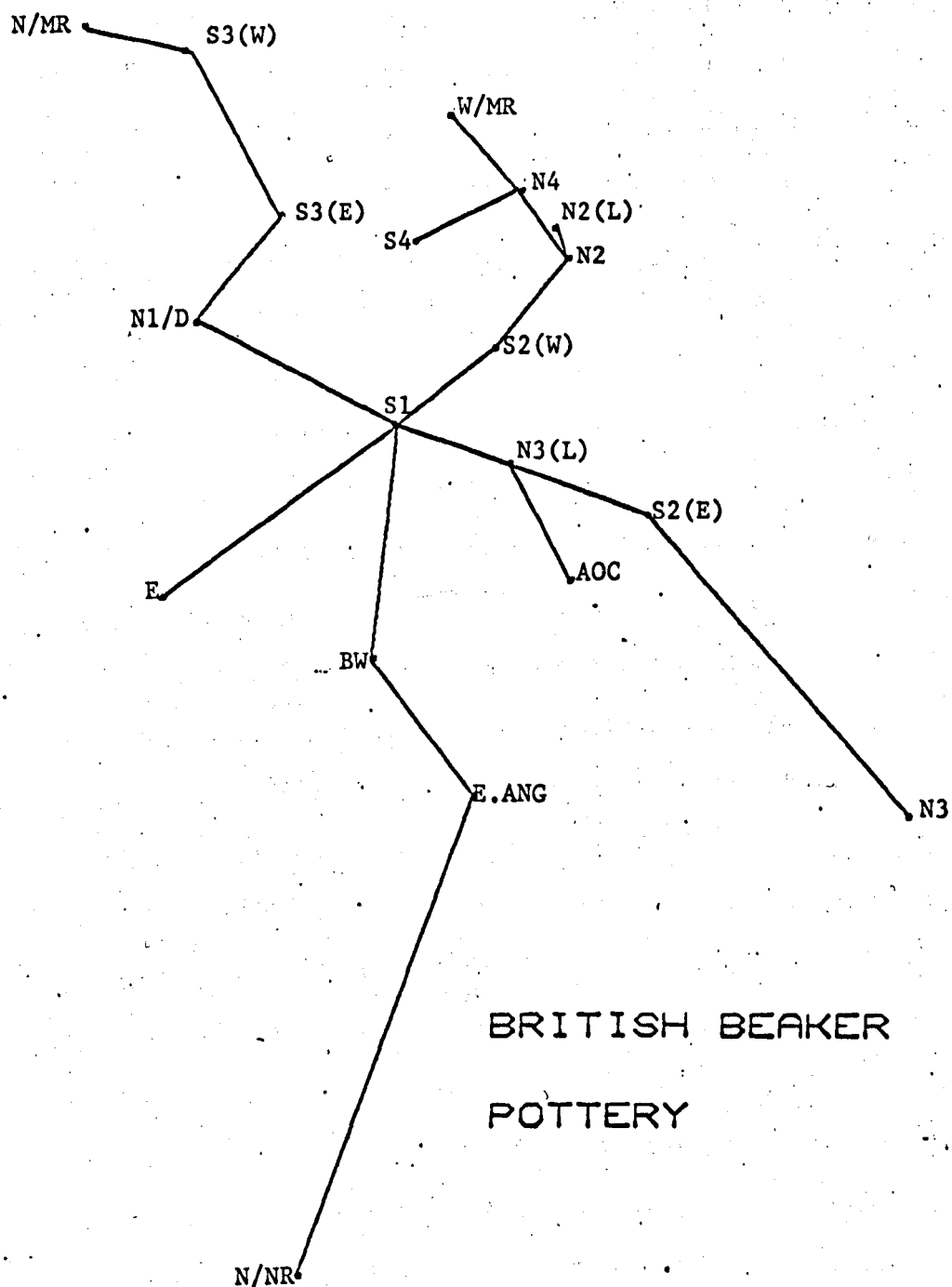


Figure 1.4

N1/D is linked between S3(E) and S1, but the fit is not good, and this would more appropriately be linked separately direct to S1. A new "intrusive" type N3 is identified (on shape alone) by the method, being squat and highly re-entrant in profile. Clarke places type N3 between N2 and N4 - this is not shown by the scalogram, and the reason may be that Clarke uses criteria other than shape (e.g. decoration) in formulating his similarity measures. The significance of the central positioning of the S1 type is that it has an "average" profile from which most others can be derived by distortions at various points. Clarke clearly identifies evolution of shape in his *Southern* series (although again S3 seems anomalous), but the analysis indicates there may be shape similarities in other directions.

- c) similarity coefficients for Eastern European projectile points. This classification is a joint project with Philip Allsworth-Jones of Emmanuel College, Cambridge. The data is a very small part of a corpus of projectile point data collected in Czechoslovakia as part of a study of Eastern European artefacts. The coefficients were generated by an original algorithm in program ARCJW028, employing Fourier analysis (see paragraph 1.10.6). The final configuration of 10 points in 2 dimensions has stress 0.125. The plain scalogram is illustrated in Figure 1.5 and with Renfrew-Sterud linkage in Figure 1.6. Although only a single projectile point from each site has been analysed, it can be seen that the method works in that points DS1, O1, JIII1, JII1 and SL1 are all medium-sized and of similar pear-shaped profile, NP1, SU1 and N10 are all small and pointed, while FD1 is elongated and MD1 very broad and blunt. The method successfully isolates the aberrant points FD1 and MD1. Again a central point

MULTIDIMENSIONAL SCALING PLOT

EASTERN EUROPEAN PROJECTILE POINTS

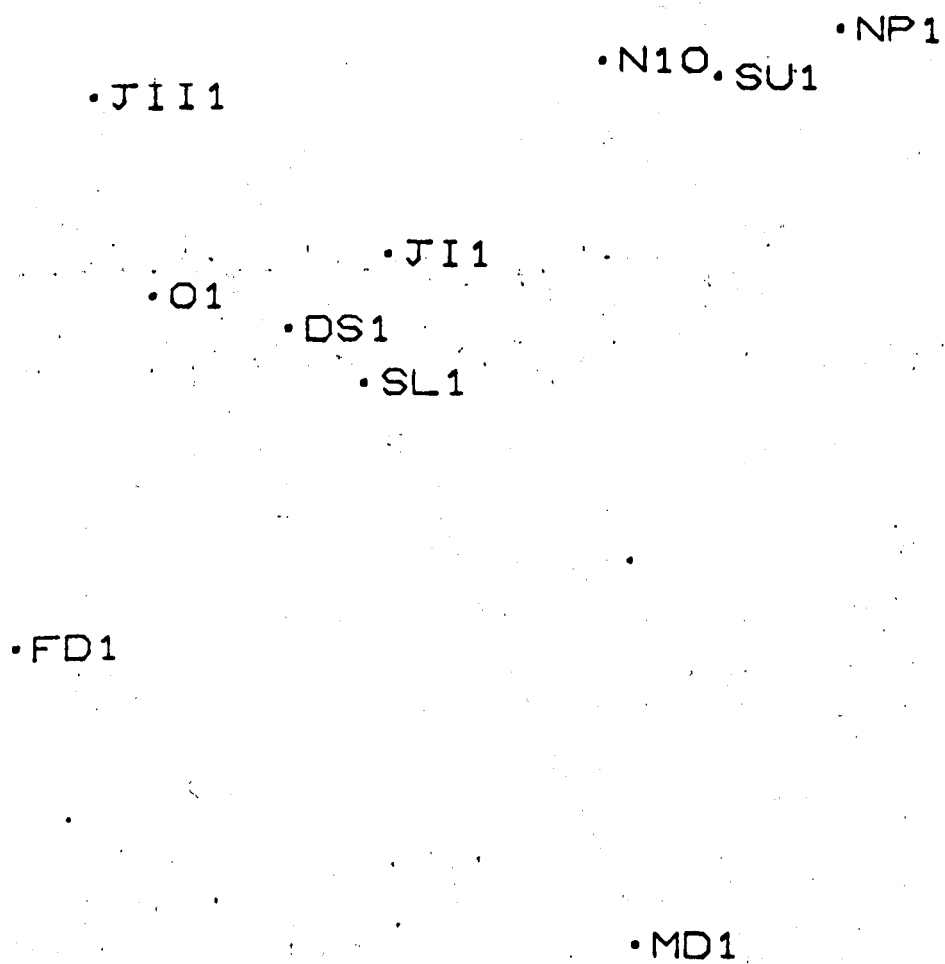


Figure 1.5

MULTIDIMENSIONAL SCALING PLOT
EASTERN EUROPEAN PROJECTILE POINTS
WITH RENFREW-STERUD LINKAGE

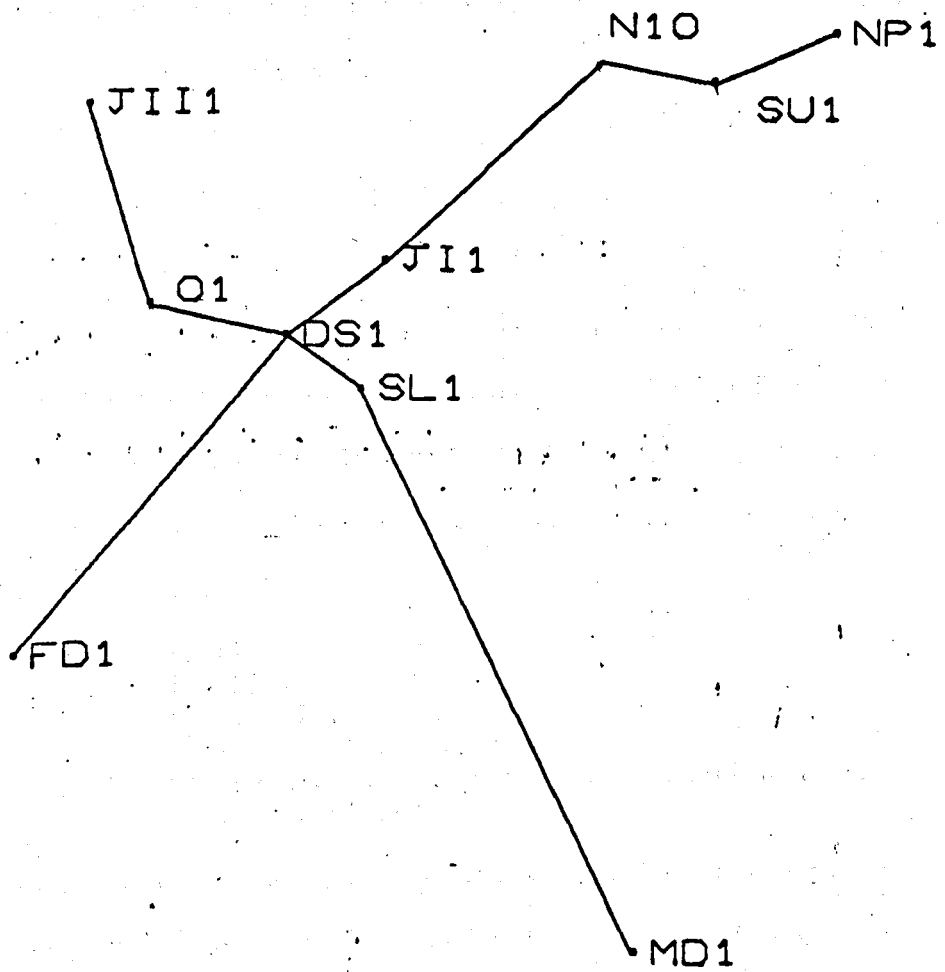


Figure 1.6

DS1 is linked to all others by the Renfrew-Sterud linkage.

DS1 has an "average" size and profile. This is a pilot study and is to be extended to cover the whole body of data when time allows. All the figures have been generated by display unit, using Graphics program ARCJWO24.

1.10.2

ARCJWO19 Pottery profile suite (ALGOL 60)

This program is used to generate similarity coefficients (based on the profiles and sizes of pots) for MDSCAL input as described in paragraph 1.10.1 above. It is fully described in the Pottery Classification chapter, paragraph 4.4.3.

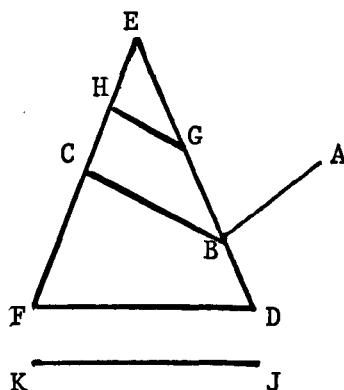
1.10.3

ARCJWO21 Masons' Marks analysis (ALGOL 60)

This program was developed to classify Masons' marks and to match similar marks using pattern recognition principles. Masons' marks tend to be simple geometrical figures formed of a few intersecting straight lines. Relatively few marks have curved portions. In order to describe these marks a series of intersection codes was developed, as follows:

0000	starting end
0001	finishing end
9000	starting intersection
9001	finishing intersection
9999	intersection which is part of a continuous line
0002	angle in a closed figure.

These codes are entered using the manual keyboard onto the paper tape carrying the d-Mac coordinates for the points of the figure which is being described. For example, the following Mason's mark



may be described as

```
0000
coordinates of A (starting end)
9999
coordinates of B (intersection on continuous line)
9001
coordinates of C (finishing intersection)
0002
coordinates of D (start at angle in closed figure)
9999
coordinates of B (intersection on continuous line)
9999
coordinates of G (intersection on continuous line)
coordinates of E
9999
coordinates of H (intersection on continuous line)
9999
coordinates of C (intersection on continuous line)
coordinates of F
0001
coordinates of D (finishing end)
9000
coordinates of G (starting intersection)
9001
coordinates of H (finishing intersection)
0000
coordinates of J (starting end)
0001
coordinates of K (finishing end)
```

Note that points C,D,G and H have more than one designation. This is because the mark is drawn by digital incremental plotter in several discrete sections (A-B-C, D-B-G-E-H-C-F-D, G-H and J-K in the example above) and the following codes 0001 and 9001 are used to lift the plotter pen, while 0000, 9000 and 0002 lower the pen. Code 9999 has no effect on the pen position. The codes are also used to identify significant points of the figure during the pattern recognition process. Note also that points E and F have no designation, since they are points on the continuous outline of the figure. Several successive points without designations may occur if a complex curve is to be defined between starting and finishing points (in this case the plotter draws the curve as a series of short straight lines). It is customary to perform a smoothing algorithm on d-Mac coordinates of complex curves to eliminate hand-shake (this is done by recalculating

the position of each successive point as the mid-point of the straight line joining its two neighbouring points), but in the case of straight-line angled figures smoothing is undesirable since the sharp angles would be eliminated (no smoothing is required in the Mason's mark above). Accordingly two extra codes are defined to control smoothing, as follows:

0003 switch off smoothing
0004 switch on smoothing

and these codes may be included anywhere in the designation of a figure. The interpretation of the codes is carried out by procedures READFIG and DRAWFIG (see Appendix A). READFIG reads the name of the Mason's mark as a string, the scale factor, then the d-Mac coordinates and codes for the figure as described above. The first two coordinates are taken as the ends of a truly vertical reference line. All the coordinates are rotated until the figure is truly vertical, smoothing is carried out where required, and the figure is scaled to a standard height with adjustment of the scale factor.

ARCJW021 has two types of run, the first a transcribing and statistics run, and the second an information retrieval run, where a "type specimen" read from paper tape is compared with all records on a magnetic tape. In the transcribing and statistics run the figure coordinates read from paper tape are searched, and the number of ends and intersections are counted separately. Points which are either starting ends or starting intersections are noted and their corresponding finishing points found. The lengths of all straight lines in the figure are recorded together with the mid-points of the lines, and used to calculate an overall centroid for the figure. The polar angles of the termination points with respect to the centroid are also recorded, and sorted into anticlockwise order. The full record

(consisting of name, coordinates including termination codes, scale factor, number of ends and intersections, coordinates of the centroid, polar angles of termination points and lengths of lines) is written to magnetic tape and is the basis for future comparison of items. In the graphics mode the program also draws and labels the figure.

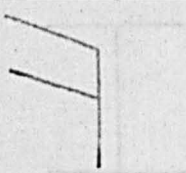
In an information retrieval run, the "type specimens" are not written to magnetic tape, but the magnetic tape is searched and a comparison made between the "type specimen" and every item on the tape. Every pair of termination points in the "type specimen" is compared with every pair of termination points in the magnetic tape record, and attempts are made to match the two figures by rotation, translation and scaling. The progress of these operations are monitored by printout. In the graphics mode the two figures are plotted in the rotated, translated and scaled transformation, and it is then possible to judge visually the degree of fit. Figure 1.7 shows the digital incremental plotter output for a typical series of Masons' marks, and Figure 1.8 the rotation translation and scaling process for a series of matches of the Roman Masons' marks WB-S 2A2 and WB-S 2A4 (see Figure 1.7). The program uses procedures SEARCH, OUTSBL, UNPACK, READFIG and DRAWFIG (see Appendix A).

1.10.4

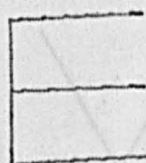
ARCJW024 Diagram Generator for Display Unit (ALGOL 60 and EDGAR)

This program is used to generate archaeological diagrams for direct publication. Two of the most common applications are the generation of phenon diagrams (dendrograms) and scalograms. Renfrew-Sterud linkage may be added to a scalogram, if desired, using the line-drawing facility. The program is fully described in the Graphics chapter, paragraph 3.4.15.

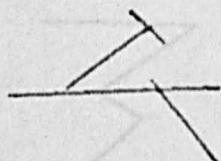
ROMAN WB-S 2A8



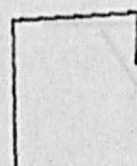
ROMAN WB-S 2A7



ROMAN WB-S 2A6



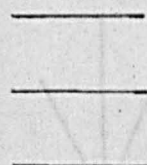
ROMAN WB-S 2A5



ROMAN WB-S 2A4



ROMAN WB-S 2A3



ROMAN WB-S 2A2



ROMAN WB-S 2A1

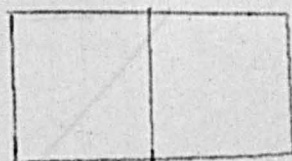
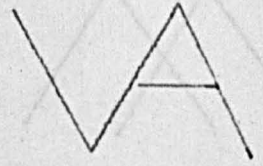


Figure 1.7
Sheet 1 of 4

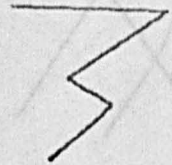
ROMAN WB-S 2A16



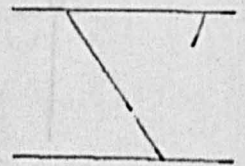
ROMAN WB-S 2A15



ROMAN WB-S 2A14



ROMAN WB-S 2A13



ROMAN WB-S 2A12



ROMAN WB-S 2A11



ROMAN WB-S 2A10



ROMAN WB-S 2A9

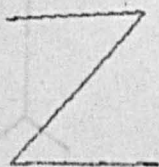
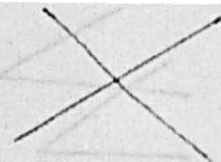


Figure 1.7
Sheet 2 of 4

CERNE ABBAS TITHE BARN WB-S 6A



SKIPTON CASTLE WB-S 280



FURNESS ABBEY WB-S 280



ROMAN WB-S 2A27



ROMAN WB-S 2A25



ROMAN WB-S 2A24



ROMAN WB-S 2A21



ROMAN WB-S 2A17

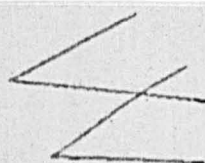


Figure 1.7

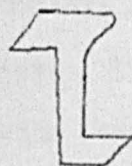
Figure 1.7

Sheet 3 of 4

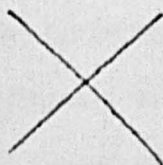
KIRKSTALL ABBEY CVB 8-28



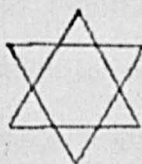
KIRKSTALL ABBEY CVB 7-8



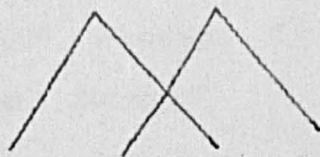
KIRKSTALL ABBEY CVB 7-6



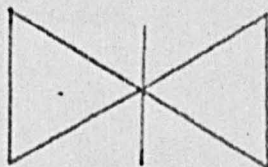
MILTON ABBEY WB-S 7A4



MILTON ABBEY WB-S 2K



LITTON CHENEY WB-S 6F



WINTERBORNE ST. MARTIN WB-S 6E

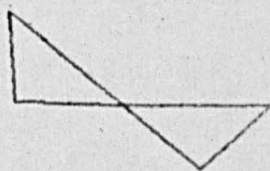
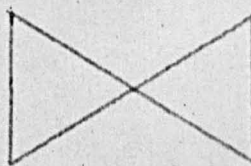


Figure 1.7
Sheet 4 of 4

MILTON ABBEY WB-S 6C



ARCJW027 Projectile Point Analysis (AL001 80 and 80000)

This program uses Fourier analysis to calculate parameters for the profiles of projectile points or any other closed figures. The profiles are produced by d-Mac and are standardised before the Fourier analysis up to order 5 is carried out. The Fourier coefficients for the y and x traces and their power series (sum of squared coefficients) are printed and also punched on paper tape for analysis by ARCJW028 (see paragraph 1.10.6). The profile of the projectile point may be synthesised from these coefficients and power spectra. It is of course found that sharp points are rounded and irregular outlines smoothed out considerably. Perfect synthesis of the closed curve would require a complete Fourier transform. Experiments have been made with up to 10 orders; in some respects (see Figure 1.9, 1.10) e.g. exaggeration of small irregularities in the profile, the results were worse than the cruder profile generated by 5 orders, and in view of the excess of similarity coefficients generated by 5 orders (see ARCJW028, paragraph 1.10.6) the extra calculation is not considered justified. ARCJW027 is described in more detail in the Graphics chapter, paragraph 3.4.18. This method of profile parameter generation using Fourier analysis applied to archaeological artefacts has not been used elsewhere.

1.10.6

ARCJW028 Projectile Point Similarity (AL001 80)

This program reads a paper tape produced by ARCJW027, consisting of series of projectile point statistics in the form:

1. Name (string enclosed in string quotes);
2. Area in sq. cm;
3. Adjusted scale factor;

Figure 1.8

ation, translation and scaling of masons' marks, a pattern recognition process

1.10.5

ARCJW027 Projectile Point Analysis (ALGOL 60 and EDGAR)

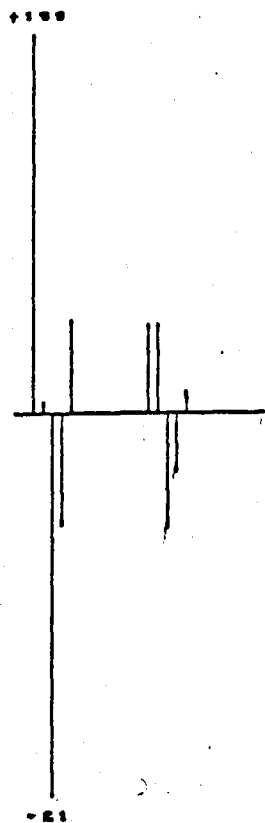
This program uses Fourier analysis to calculate parameters for the profiles of projectile points or any other closed figures. The profiles are produced by d-Mac and are standardised before the Fourier analysis up to order 5 is carried out. The Fourier coefficients for the y and x traces and their power series (sum of squared coefficients) are printed and also punched on paper tape for analysis by ARCJW028 (see paragraph 1.10.6). The profile of the projectile point may be synthesized from these coefficients and power spectra. It is of course found that sharp points are rounded and irregular outlines smoothed out considerably. Perfect synthesis of the closed curve would require a complete Fourier transform. Experiments have been made with up to 10 orders; in some respects (see Figures 1.9, 1.10) e.g. exaggeration of small irregularities in the profile, the results were worse than the cruder profile generated by 5 orders, and in view of the success of similarity coefficients generated by 5 orders (see ARCJW028, paragraph 1.10.6) the extra calculation is not considered justified. ARCJW027 is described in more detail in the Graphics chapter, paragraph 3.4.18. This method of profile parameter generation using Fourier analysis applied to archaeological artefacts has not been seen elsewhere.

1.10.6

ARCJW028 Projectile Point Similarity (ALGOL 60)

This program reads a paper tape produced by ARCJW027, consisting of series of projectile point statistics in the form:

1. Name (string enclosed in string quotes);
2. Area in sq. cm;
3. Adjusted scale factor;
4. Fourier transform coefficients and power series for five orders.



DZERAVA SKALA 1

SCALE 1.32

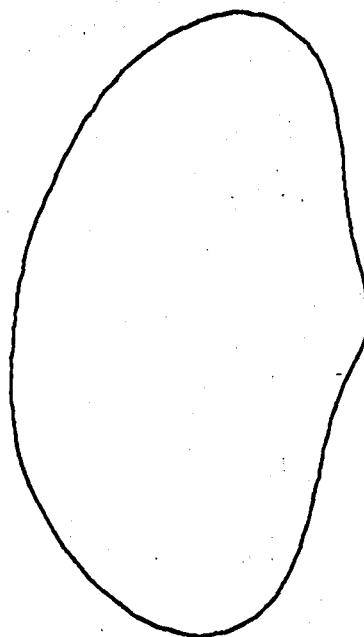
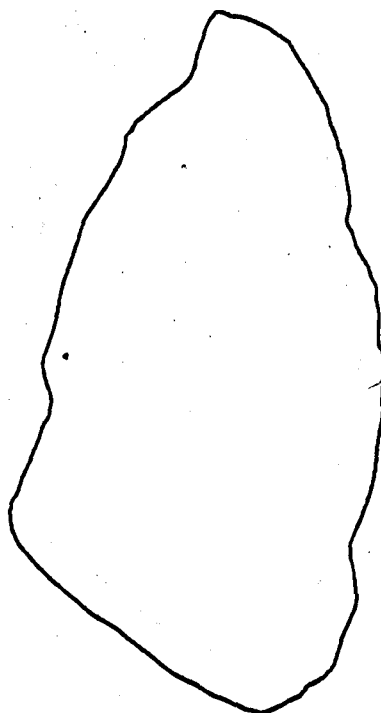
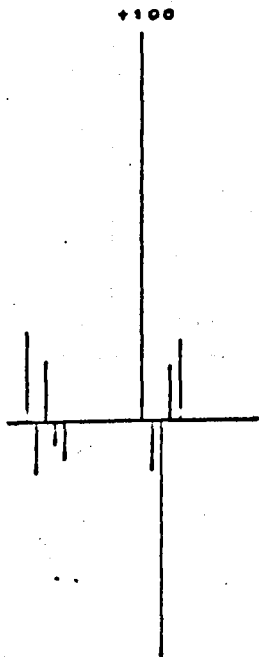
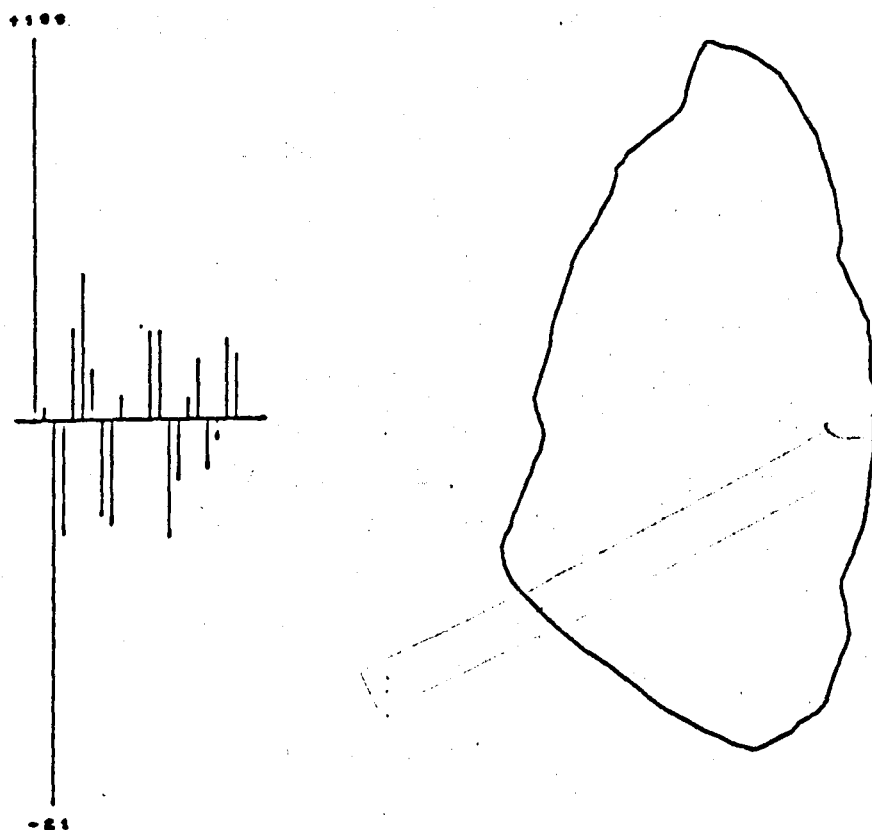


Figure 1.9



DZERAVA SKALA 1

SCALE 1.32

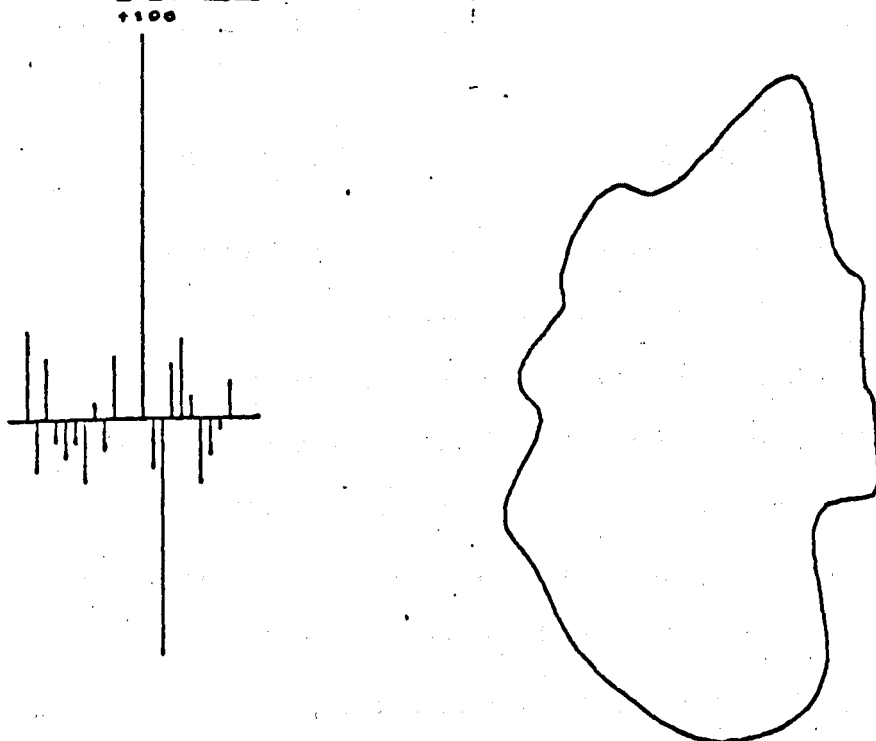


Figure 1.10

Fourier analysis of projectile points - 10 orders

The reading phase continues until an area < 0.5 sq.cm is read (this is automatically produced when ARCJW027 is terminated).

Each record is then compared with every other record and similarity coefficients are calculated in half-matrix, diagonal-absent format as required by ARCJW018 (MDSCAL). Each record is also compared with itself as a check. A similarity coefficient based on Fourier analysis alone (S_f) is defined as:

$$S_{f_{ij}} = 100 - \sum_{k=1}^5 \left\{ \left(\left(a_{y_{ki}} - a_{y_{kj}} \right)^2 + \left(b_{y_{ki}} - b_{y_{kj}} \right)^2 \right) / p_{y_{ki}} p_{y_{kj}} + \left(\left(a_{x_{ki}} - a_{x_{kj}} \right)^2 + \left(b_{x_{ki}} - b_{x_{kj}} \right)^2 \right) / p_{x_{ki}} p_{x_{kj}} \right\}$$

where a_{y_k} , a_{x_k} are the k^{th} order sine term coefficients for the y and x traces

b_{y_k} , b_{x_k} " " k^{th} " cosine " " " " " " "

p_{y_k} , p_{x_k} " " k^{th} " power series " " " " " "

$$\text{i.e. } p_{y_k} = a_{y_k}^2 + b_{y_k}^2$$

$$p_{x_k} = a_{x_k}^2 + b_{x_k}^2$$

Division by the geometric mean of the power series is included to normalise the Euclidian squared distances between corresponding coefficients of the two items being compared, so that the dominant magnitudes of the first order sine coefficient for the y trace and the first order cosine coefficient for the x trace do not swamp the higher order coefficients. Indeed as much, if not more importance is to be attached to the higher order coefficients, since these express the differences of the profile from the basic ellipsoidal shape which would result if only the first order coefficients were considered.

An areal similarity coefficient is defined as:

$$S_a = \frac{100 A_j}{A_i} \quad \text{if } A_i > A_j$$

$$\text{or } S_a = \frac{100 A_i}{A_j} \quad \text{if } A_i \leq A_j$$

A scale similarity coefficient is defined as:

$$S_s = \frac{100 S_j}{S_i} \quad \text{if } S_i > S_j$$

$$\text{or } S_s = \frac{100 S_i}{S_j} \quad \text{if } S_i \leq S_j$$

Using a similar argument to that employed for pottery profile similarities, the Fourier similarity S_f may be expected to carry some information about the area as well as the profile of the projectile point, while the scale similarity S_s carries information about the length of the projectile point. An overall similarity coefficient is therefore defined as

$$S = \frac{S_f S_s}{100}$$

These values are printed in half-matrix form, each coefficient being identified with the names of the two items being compared. The values are used for input to ARCJW018 (MDSCAL) with successful results. A typical printout of the similarity coefficients is given in Figure 1.11. The above method of similarity calculation gives good results (see MDSCAL output for a sample set of projectile points, Figures 1.5 and 1.6) and is claimed to be original.

1.10.7

HISTOG Histogram-plotting program

(BASIC, designed for remote terminal use)

This is a "lazy man's" histogram-plotting program, which analyses data without the user having to specify the number of items, range of values, or interval.

JEZERANY I 1

13.53716	86.46284	70.80103	96.02389	61.21659	83.02498	DZERAVA SKALA 1
21.14555	78.85445	88.16537	81.74783	69.52232	64.46180	FOURNEAU-DU-DIABLE 1
0.00000	100.00000	100.00000	100.00000	100.00000	100.00000	JEZERANY I 1

JEZERANY II 1

22.64798	77.35202	75.83942	92.27969	58.66332	71.38020	DZERAVA SKALA 1
21.29241	78.70759	60.90270	72.43719	47.93504	57.01357	FOURNEAU-DU-DIABLE 1
23.41630	76.58370	53.69509	88.61054	41.12169	67.86124	JEZERANY I 1
0.00000	100.00000	100.00000	100.00000	100.00000	100.00000	JEZERANY II 1

MORAVANY-DLHA 1

45.95905	54.04095	56.28595	88.81854	30.41746	47.99838	DZERAVA SKALA 1
38.47907	61.52093	70.09039	88.37957	43.12026	54.37194	FOURNEAU-DU-DIABLE 1
58.25144	41.74856	79.49877	92.49630	33.18959	38.61587	JEZERANY I 1
31.02914	68.97086	42.68694	81.96147	29.44155	56.52953	JEZERANY II 1
0.00000	100.00000	100.00000	100.00000	100.00000	100.00000	MORAVANY-DLHA 1

NESLOVICE 10

10.05270	89.94730	98.24818	79.83467	88.37158	71.80913	DZERAVA SKALA 1
16.19553	83.80447	78.89801	62.66817	66.12006	52.51873	FOURNEAU-DU-DIABLE 1
8.10040	91.89960	69.56072	76.66036	63.92602	70.45056	JEZERANY I 1
18.61163	81.38837	77.19168	86.51381	62.82505	70.41218	JEZERANY II 1
42.88053	57.11947	55.29992	70.90799	31.58702	40.50226	MORAVANY-DLHA 1
0.00000	100.00000	100.00000	100.00000	100.00000	100.00000	NESLOVICE 10

NESLOVICE-PRAGUE 1

19.98609	80.01391	85.32847	79.67541	68.27464	63.75141	DZERAVA SKALA 1
38.97707	61.02293	68.52286	62.54316	41.81466	38.16567	FOURNEAU-DU-DIABLE 1
4.49485	95.50515	60.41344	76.50743	57.69794	73.06853	JEZERANY I 1
42.60680	57.39320	88.87938	86.34122	51.01072	49.55399	JEZERANY II 1
88.78373	11.21627	48.02794	70.76654	5.38694	7.93736	MORAVANY-DLHA 1
14.12313	85.87687	86.84993	99.80051	74.58400	85.70556	NESLOVICE 10

Numerical quantities are,
in order left to right:
Fourier difference coefficient
FOURA,
SIM = 100 - FOURA,
Area ratio AREAR,
Scale ratio SCR,
SIM*AREAR/100,
and SIM*SCR/100.

Figure 1.11
Sheet 1 of 2

Figure 1.11
Sheet 1 of 2

0.00000 100.00000 100.00000 100.00000 100.00000 100.00000 NESLOVICE-PRAGUE 1

ONDRATRICE 1

18.27275	81.72725	97.64790	96.99154	79.80494	79.26852	DZERAVA SKALA 1
20.58168	79.41832	82.23916	80.93225	65.31296	64.27504	FOURNEAU-DU-DIABLE 1
19.96525	80.03475	72.50646	99.00233	58.03037	79.23627	JEZERANY I 1
9.25719	90.74281	74.05560	89.50349	67.20013	81.21798	JEZERANY II 1
31.25492	68.74508	57.64174	91.57349	39.62586	62.95227	MORAVANY-DLHA 1
14.80705	85.19295	95.93728	77.43288	81.73180	65.96736	NESLOVICE 10
37.19987	62.80013	83.32145	77.27841	52.32598	48.53094	NESLOVICE-PRAGUE 1
0.00000	100.00000	100.00000	100.00000	100.00000	100.00000	ONDRATRICE 1

Numerical quantities are,
in order left to right:
Fourier difference coefficient
FOURA,
SIM = 100 - FOURA,
Area ratio AREAR,
Scale ratio SCR,
SIM*AREAR/100,
and SIM*SCR/100.

SZELETA-LOWER 1

8.50422	91.49578	97.30114	96.06401	89.02644	87.89452	DZERAVA SKALA 1
12.45664	87.54336	82.53224	75.40779	72.25150	66.01451	FOURNEAU-DU-DIABLE 1
12.10091	87.89909	72.76486	92.24439	63.95965	81.08198	JEZERANY I 1
19.12545	80.87455	73.79261	96.06063	59.67945	77.68860	JEZERANY II 1
27.83138	72.16862	57.84717	85.32265	41.74750	61.57617	MORAVANY-DLHA 1
14.27254	85.72746	95.59659	83.10571	81.95253	71.24441	NESLOVICE 10
24.53255	75.46745	83.02557	82.93992	62.65728	62.59265	NESLOVICE-PRAGUE 1
15.22774	84.77226	99.64489	93.17396	84.47122	78.98567	ONDRATRICE 1
0.00000	100.00000	100.00000	100.00000	100.00000	100.00000	SZELETA-LOWER 1

SZELETA-UPPER 1

12.30180	87.69820	53.13869	75.04171	46.60167	65.81023	DZERAVA SKALA 1
15.50810	84.49190	42.67292	58.90582	36.05516	49.77065	FOURNEAU-DU-DIABLE 1
4.94243	95.05757	37.62274	72.05797	35.76326	68.49656	JEZERANY I 1
25.33234	74.66766	70.06737	81.31986	52.31767	60.71964	JEZERANY II 1
64.74908	35.25092	29.90961	66.65095	10.54342	23.49508	MORAVANY-DLHA 1
11.54201	88.45799	54.08618	93.99639	47.84355	83.14732	NESLOVICE 10
15.39486	84.60514	62.27545	94.18428	52.68823	79.68474	NESLOVICE-PRAGUE 1
21.96381	78.03619	51.88881	72.78411	40.49205	56.79795	ONDRATRICE 1
8.80621	91.19379	51.70455	78.11637	47.15133	71.23727	SZELETA-LOWER 1
0.00000	100.00000	100.00000	100.00000	100.00000	100.00000	SZELETA-UPPER 1

Figure 1.11

Sheet 2 of 2

A string identifier is read to name the distribution of items, then printed, prefixed by DISTRIBUTION. Numerical items are then read from the data list until a terminator is detected. The items are counted, and the count is printed. The maximum and minimum values are also printed, followed by the range.

The interval for the histogram is calculated automatically, as follows:

$$l = \text{INT}(\log_{10} \text{range}) \quad \text{where INT is the lower integer}$$

$$S = \text{range} / 10^l$$

Value of S	Interval chosen automatically
------------	-------------------------------

$1 \leq S < 2$	$10^{(l-1)}$
----------------	--------------

$2 \leq S < 5$	$2 \times 10^{(l-1)}$
----------------	-----------------------

$5 \leq S < 10$	$5 \times 10^{(l-1)}$
-----------------	-----------------------

and the interval is printed. The items are allocated to the histogram blocks and a count produced for each. The histogram is then printed on the teletype, each histogram being identified by its central value, number of items allocated to the block, and a symbolic representation of this number. If the number of symbols required would cause the carriage width to be exceeded, a default symbol is output instead.

Typical runs are shown in Figures 1.12 and 1.13. Figure 1.12 also shows the data entered by the archaeologist before the run. Note that there is no need to count the items, find the range or decide on a suitable interval. A program giving such facilities for archaeologists has not been seen elsewhere.

1.10.8

LIFE. A simulation and model investigation program (BASIC, designed for remote terminal use)

This is a simulation and model investigation program which sets

2000 DATA 102,102,102,103,110,96,110,109,103,102,103,108,115,94,102
 2010 DATA 101,102,101,95,101,98,103,105,97,112,98,98.0,95,106,100,106
 2020 DATA 101,105,105,105,110,107,99,104,101,99,104,110,109,106,107
 2030 DATA 100,106,98,99,94,105,103,105,113,105,109,107,107,106,102,102
 2040 DATA 107,112,110,103,103,105,110,108,99,100,105,105,104,107,103
 2050 DATA 110,105,106,106,104,105,100,97,103,104,96,110,104,101,107
 2060 DATA 92,106,106,98,94,111,102,96,91,92,95,105,96,99,108,102,103
 2070 DATA 108,103,99,95,100,95,97,102,109,104,99,97,105,107,109,101
 2080 DATA 100,110,108,104,106,10000

SAVE

RUN
 DISTRIBUTION 148 STEMS
 NO. ITEMS 130
 MAXIMUM 115
 MINIMUM 91
 RANGE 24
 STEP 2

90	0	
92	0	***
94	3	***
96	3	*****
98	9	*****
100	9	*****
102	13	*****
104	18	*****
106	19	*****
108	24	*****
110	13	*****
112	14	*****
114	3	***
	1	*
	1	*

READY

Figure 1.12

HISTOG run with data for clay pipe bores

NO. ITEMS	60
MAXIMUM	83.4
MINIMUM	-99.6
RANGE	183.
STEP	10

-100	0	
-90	1	*
-80	0	
-70	0	
-60	0	
-50	0	
-40	0	
-30	0	
-20	0	
-10	0	
0	28	*****
10	10	*****
20	1	*
30	2	**
40	11	*****
50	6	*****
60	0	
70	0	
80	1	*
	0	

READY

Figure 1.13

HISTOG run

up a model for the growth and interaction of an archaeological culture. The illustrated model is fairly simple but the program could easily be elaborated to carry out simulations with complex control parameters. In this example the criteria are that during each time period:

- a) a unit may expand in a random direction by one unit provided at least one adjacent site is vacant;
- b) when all possible expansion has taken place, units which are now overcrowded (i.e. surrounded by 4 neighbour units on adjacent sites) die and are deleted;
- c) when all necessary deletion has taken place, units which are now isolated (i.e. have no neighbour units) also die and are deleted.

The program clears an area of store to represent a finite area. A border of pseudo-units (value 10) is also set up to prevent edge effect caused by the finite area, i.e. non-standard conditions arising because a unit at the edge cannot be overcrowded. In a truly infinite area such a unit could be overcrowded. The border elements are, however, ignored in the case of a check for isolation, again to prevent non-standard conditions.

The number of time periods to be simulated is read, followed by the location of units in the initial configuration. Units are set up (value 1) in the area until a terminator is encountered. The title INITIAL DATA is printed, then a subroutine is entered to print out the area,* being used to indicate the presence of a unit. A similar printout occurs for each time period, being headed by the title PERIOD n.

Entering the expansion cycle, for every unit which is present a random integer between 1 and 4 is generated and used to indicate the direction of the adjacent square to be chosen for expansion.

If this square is vacant, a new unit (value 2) is set up there and the next of the *original* units is examined for expansion (the program discriminates between first and second-generation units). If the chosen square is already occupied by a first or second-generation unit, up to a further nine random numbers may be generated in an attempt to expand the current first-generation unit. The cycle repeats until all first-generation units have been examined.

A subroutine is now entered to make all units into first-generation units (value 1) then the overcrowding cycle is entered. For each unit the sum of its four nearest neighbours is calculated. These may be vacant (value 0), other unmarked units (value 1), marked units (value 2) or border pseudo-units (value 10). If the sum is < 4 the unit cannot be a candidate for overcrowding. If $4 \leq \text{sum} \leq 6$ the unit might be surrounded by all unmarked units (sum 4), or a mixture with up to two marked units (since marked units can only occur on two sides because of the processing method), or there may only be three or two adjacent units - the unit is checked to have four adjacent units and is then marked with value 2 for deletion. Similar arguments apply for sums in the ranges 13 to 15 and 22 to 24 (see the table below).

<i>Sum of the four nearest neighbours</i>	<i>Possible values for four nearest neighbours</i>	<i>Deletion (Indicated by *)</i>	<i>Remarks</i>
0	0 0 0 0		
1	0 0 0 1		
2	0 0 1 1		
	0 0 0 2		
3	0 1 1 1		
	0 0 1 2		
4	1 1 1 1	*	
	0 1 1 2		
	0 0 2 2		
5	1 1 1 2	*	
	0 1 2 2		
6	1 1 2 2	*	Only two 2 units possible
10	0 0 0 10		1 border unit
11	0 0 1 10		
12	0 1 1 10		
	0 0 2 10		
13	1 1 1 10	*	
	0 1 2 10		
14	1 1 2 10	*	
	0 2 2 10		
15	1 2 2 10	*	
20	0 0 10 10		2 border units
21	0 1 10 10		
22	1 1 10 10	*	
	0 2 10 10		
23	1 2 10 10	*	
24	2 2 10 10	*	Only two border units possible

At the end of this pass the overcrowded units are deleted (they were not deleted immediately since they were still needed to determine the possible overcrowding of subsequent units).

The isolation cycle is now entered. A sum is formed similarly for the four adjacent units. If it is zero, the unit is immediately deleted, and this is also done if one or two border elements only are adjacent.

The printout subroutine is called to print out the result. After the required number of time periods have been executed, a further set of initial data is read or the program terminates.

Typical runs of the program are given in Figures 1.14, 1.15 and 1.16. Figures 1.14 and 1.15 are for the same initial configuration, but the random nature of the expansion produces different results. Figure 1.16 is for a second initial configuration.

1.10.9

PHENON Dendrogram program

(BASIC, designed for remote terminal use)

This program sorts items into the order required for a dendrogram, using an average-link clustering strategy. A string variable to identify the data is read and printed. The half-matrix less diagonal representing the similarity coefficients for each item against every other item is read from the data list. A check is incorporated to detect rows of incorrect length, with failure message

DATA REGISTRATION ERROR LINE n

The discrimination level is set to 100%. The initial number of clusters is set equal to the number of items (i.e. columns in the matrix). Each cluster is allocated a single item and the clusters are numbered sequentially.

Entering the main loop of the program, each existing cluster is

RUN

INITIAL DATA

PERIOD

5

PERIOD

1010

*

```
*** *
***
** *
* *
* **
**
*
```

```
  * *
*** * *
*      **
***      *
      **
*   ** *
*   * *
*   * *
*   * *
```

PERIOD

1

PERIOD

6

*
*

```
** **
*   *
*   *
*   **
*   ***
*   * *
**
*
```

PERIOD

2

PERIOD

7

*
*
**

```
**
* *****
** **
** **
*   * *
*** * *
*   *
*   ***
**
```

PERIOD

3

PERIOD

8

**
**
**
**

```
*
***
* *****
*   ** *
*       *
*   ** *
*   *** *
**   * *
*   ***
*
```

PERIOD

4

PERIOD

9

*

* *
* *

*

```
*
***** *
*** * **
* * **
** **
* **
** * **
**   **
***   *
*     *
```

Culture growth
model with same
initial configuration
as Figure 1.14,
showing different
random directions of
Culture
growth
model

Figure 1.15

Figure 1.14

INITIAL DATA

PERIOD

5

PERIOD

10

*

```

**
* ****
**  *
      **
      **
    **
  
```

```

  *  *
***** **
      *** *
    *  * **
    ***
** ** **
      *
      ** *
** ** *
    ** *
  
```

PERIOD

1

PERIOD

6

**

```

**
** **
* ** *
***** * *
      **
      ***
    **
  
```

PERIOD

2

PERIOD

7

**
**

```

  *
***  *
* *****
*      * *
*      **
** * *
  * *
  **
  
```

PERIOD

3

PERIOD

8

*

```

  **
*** * ***
****  *
      * *
    * * *
***** * *
      ** *
    **** *
  
```

PERIOD

4

PERIOD

9

*** **
** *

```

*****
      *****
      **
    * **
    ** ***
***  * *
** * **
*** ***
    *
  
```

Culture growth model with same initial configuration as Figure 1.14, showing different random directions of expansion

Figure 1.15

PERIOD

1

PERIOD

6

READY

PERIOD

2

PERIOD

7

PERIOD

3

PERIOD

8

PERIOD

4

PERIOD

9

Culture growth model with an initial configuration which is different from that used in the simulations illustrated in Figures 1.14 and 1.15

Figure 1.16

examined in turn. If the similarity between the cluster under consideration and another cluster is \geq the discrimination level, the two clusters are amalgamated, retaining the small identification number of the pair. The cluster of larger identification number is deleted. Similarity coefficients between the new cluster and all other existing clusters are adjusted using the average-link weighted mean:

$$S_{ik} = (n_i s_{ik} + n_j s_{jk}) / (n_i + n_j)$$

where i is the retained cluster number, j is the deleted cluster, n_i is the number of items in cluster i , and S_{ik} is the similarity coefficient between clusters i and k . The current value of the discrimination level and the identification numbers of the two amalgamated clusters are printed. The positions of the two clusters in the present order are located, and the items in the cluster to be deleted are deposited in a buffer, unless the two clusters are already adjacent; intervening items are moved into the space originally occupied by the deleted cluster, and the items of the deleted cluster are added to the retained cluster. The number of items in the retained cluster is increased by the number of items in the deleted cluster, and the number of items in the deleted cluster is set to zero to indicate that this cluster no longer exists. The new order of the items is printed.

If only one cluster now remains a further set of data is read or if there is no more data the program is terminated. Otherwise the search continues for similarity coefficients \geq discrimination level (several amalgamations may occur at any one discrimination level). If no more are found, the discrimination level is reduced by 1% and the main loop repeats.

The process is illustrated for a small set of data in the following table, which shows half-matrices of similarity coefficients in the range 0.0 - 100.0

Order

Initial data:

83					1 2 3 4 5
82	80				
96	74	76			
43	16	53	25		

Discrimination level 96:

78.5					(1 4) 2 3 5
79	80				
0	0	0			
34	16	53	0		

Discrimination level 80:

78.75					(1 4) (2 3) 5
0	0				
0	0	0			
34	34.5	0	0		

Discrimination level 78:

0					(1 4 2 3) 5
0	0				
0	0	0			
34.25	0	0	0		

Discrimination level 34:

0					(1 4 2 3 5)
0	0				
0	0	0			
0	0	0	0		

The output for a typical run, in this case for British Beaker pottery based on Clarke's classification data, is shown in Figure 1.17. The groupings are showed encircled as they are formed. A dendrogram produced by the diagram generator ARCJW024 (see paragraph 3.4.15) for the same data is shown in Figure 1.18, and a scalogram with both Renfrew-Sterud linkage and isopleths for the same data in Figure 1.19. The advantages to be gained from conjunction of the scalogram, Renfrew-Sterud linkage and isopleths, which has not been seen elsewhere, are evident. Further dendrograms generated by ARCJW024 for common Samian pottery types and Eastern European projectile points are shown in Figures 1.20 and 1.21 respectively. The results will now be discussed from the point of view of success in archaeological classification.

(a) The common Samian pottery types (see Figure 1.20) show similar groupings to those identified by multidimensional scaling (see paragraph 1.10.1). The small "dish-type" group (27,33,31,18/31,18,36,35) is formed at discrimination level 68%. Examining the evolutionary sequence 18,18/31,31 it is seen that 18/31 and 31 are grouped together at 89% similarity while 18 first joins a group with 35 and 36 at 82%, not joining 18/31 and 31 until the 68% phenon. This premature allocation of items early in the classification process to groups which are not necessarily the best for them in the final classification is a disadvantage of the *average-link* method. Allocation is irrevocable, i.e. subsequent migration is not allowed as it is in the *k-means* method. The large "bowl-type" group (37,29,45,38) is formed at 75% similarity, and is therefore a more coherent group than the "dish types". The discordant forms 30 and 68 form a group of their own at 81%, but do not join the large "bowl type" group until the low figure of 45% similarity. The whole corpus becomes one group at 33% similarity.

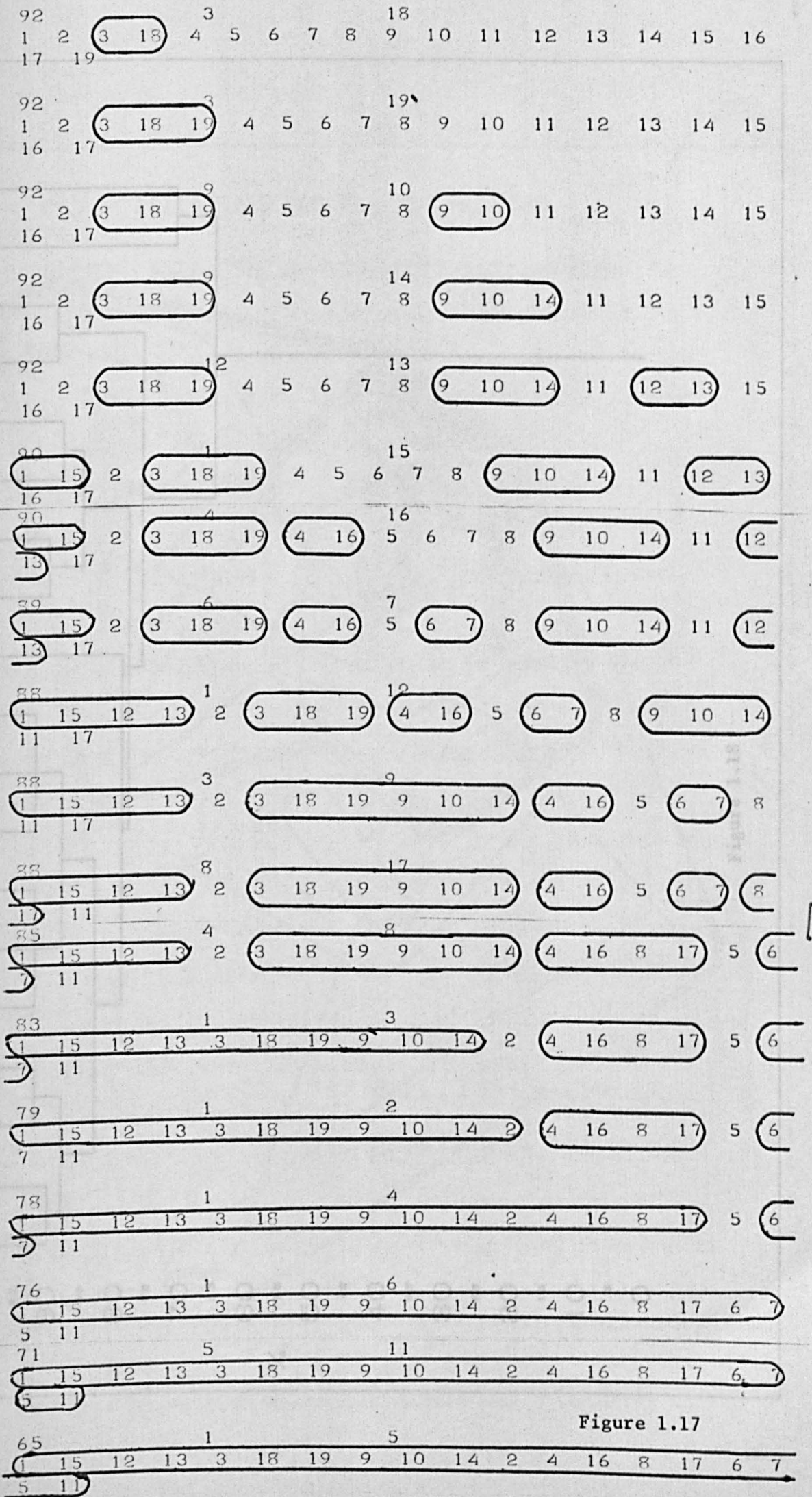


Figure 1.17

DENDROGRAM FOR BRITISH BEAKER POTTERY

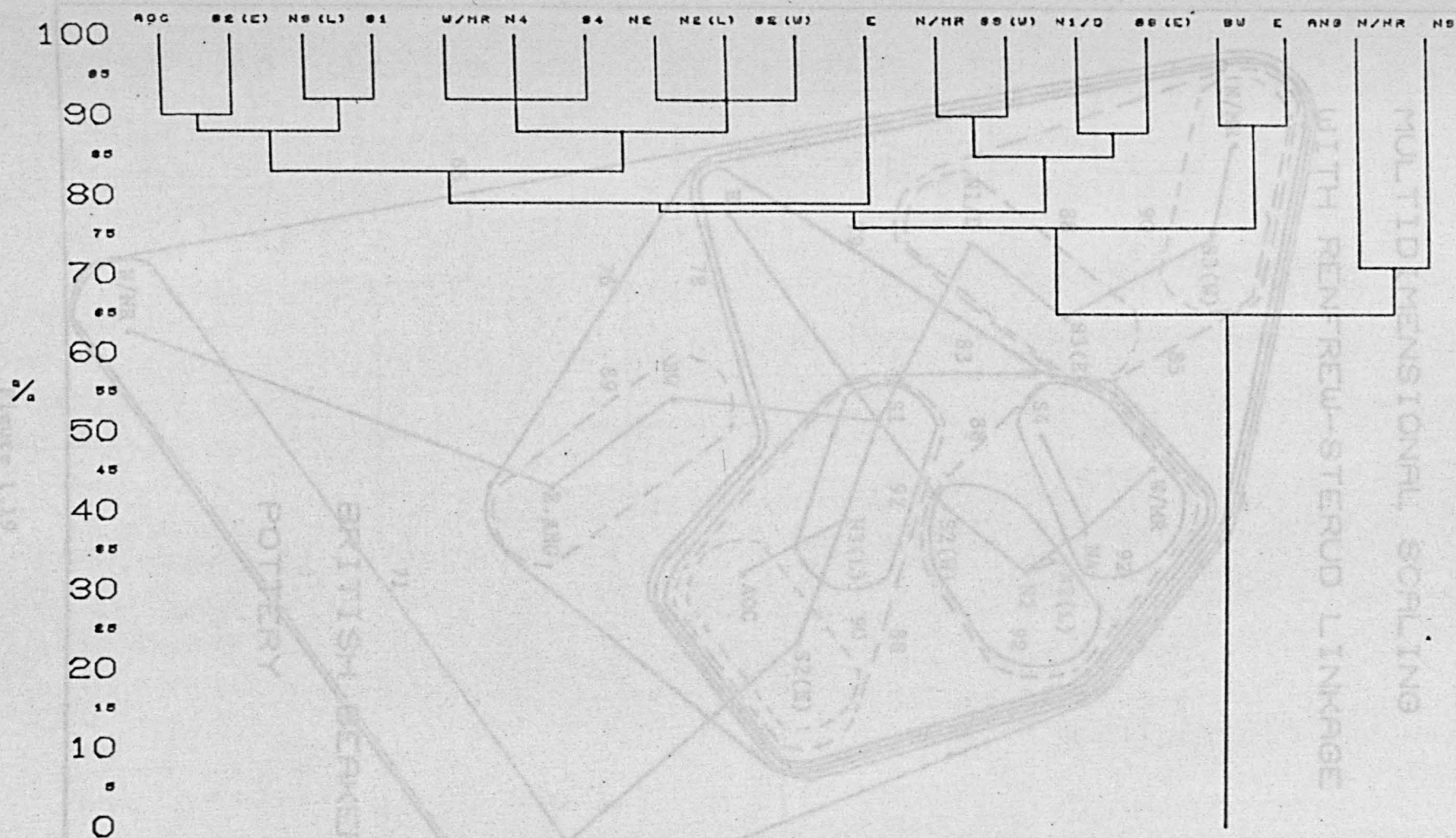


Figure 1.18

MULTIDIMENSIONAL SCALING

WITH RENFREW-STERUD LINKAGE

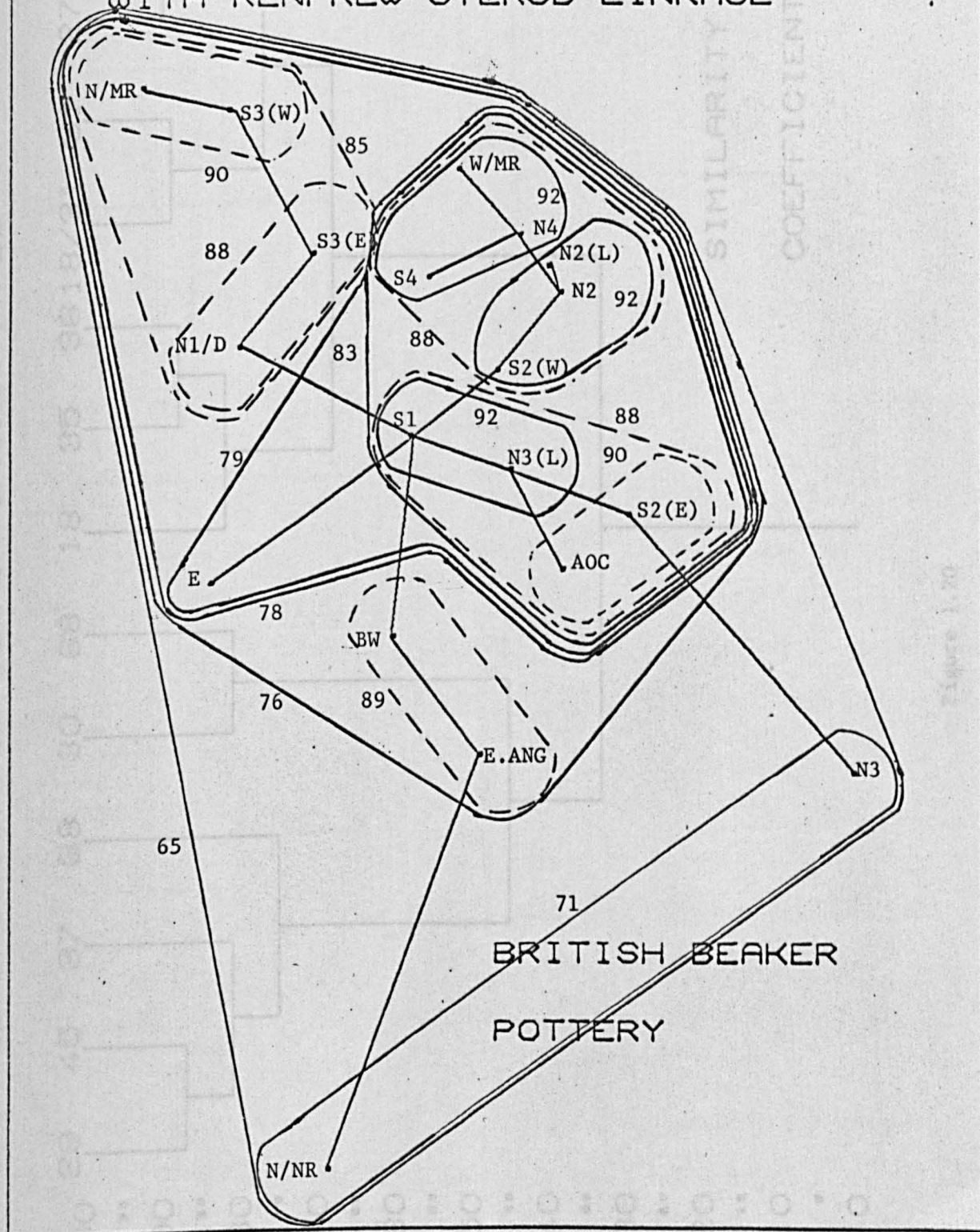


Figure 1.19
Scalogram for British Beaker pottery with phenon isopleths

PHENON DENDROGRAM FOR COMMON SAMIAN TYPES

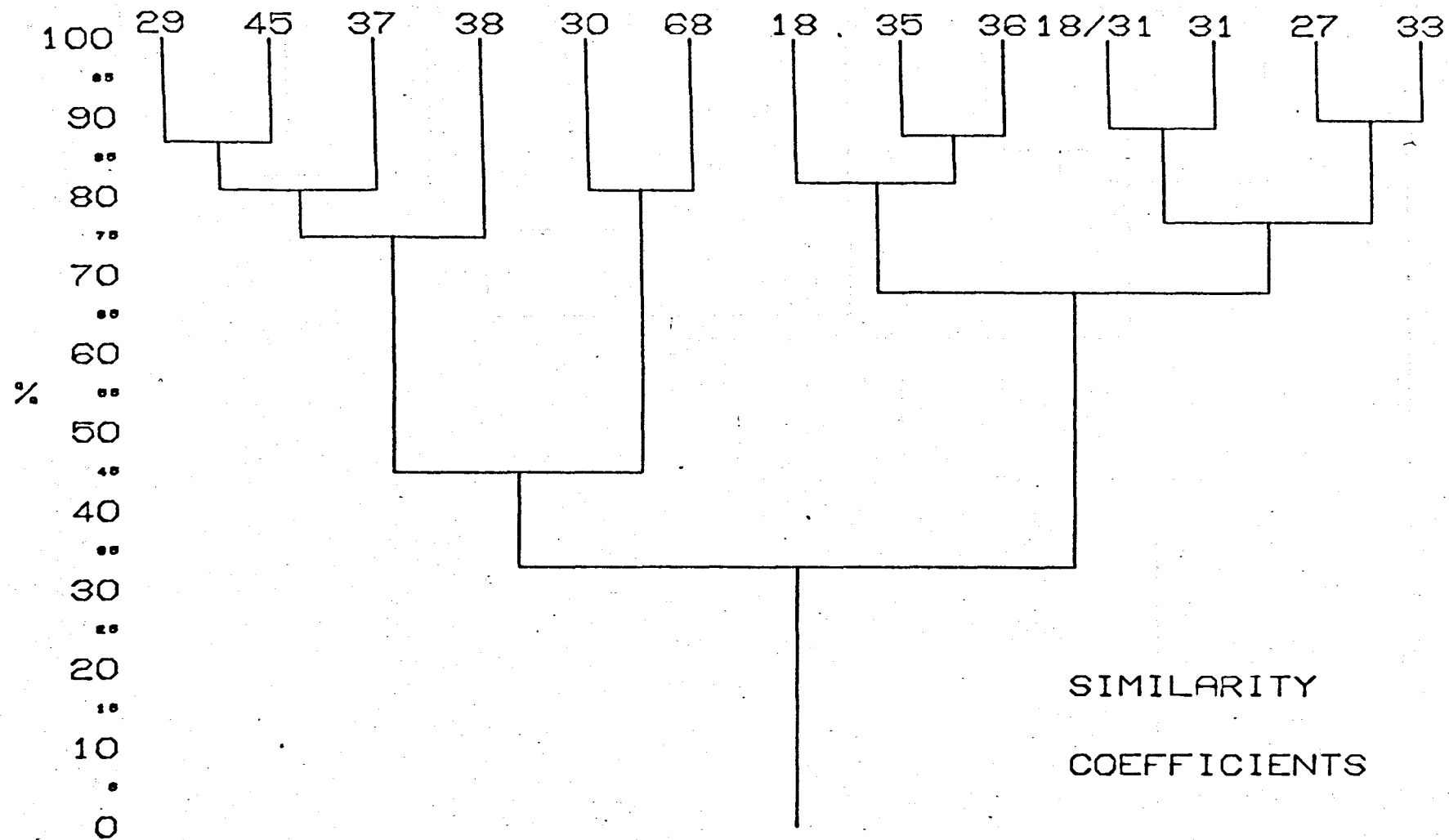


Figure 1.20

DENDROGRAM FOR EASTERN EUROPEAN PROJECTILE POINTS

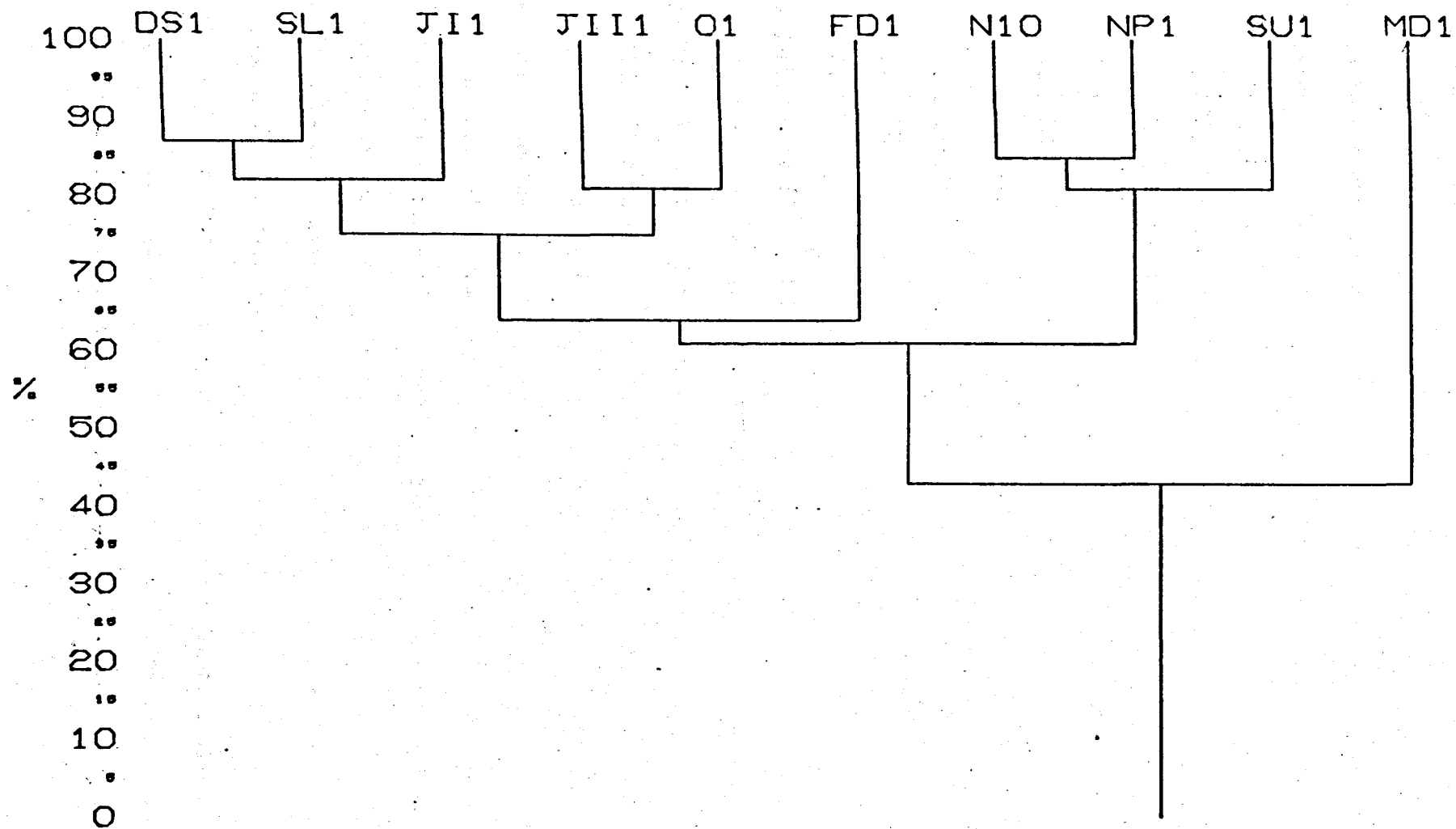


Figure 1.21

(b) The British Beaker pottery (see Figures 1.18 and 1.19) is much more similar overall, forming a single group at 65% similarity. The successive groupings are shown graphically by the isopleth method (see Figure 1.19). The main groups (based on shape and size alone) seem to be essentially defined by the 87% phenon. One of Clarke's "intrusive" types is associated with most of these groups. The groups are as follows (intrusive types in gothic script):

1. AOC, S2E, N3L, S1
2. W/MR, N4, S4, N2, N2L, S2W
3. E
4. N/MR, S3W
5. N1/D, S3E
6. BW, E.Ang.
7. N/NR
8. N3

The Barbed Wire (BW) and East Anglian (E.Ang.) intrusive types are very similar in shape, differing only in decoration. A new "intrusive" type N3 is identified (on shape alone). It is evident that this is very dissimilar in shape to types N2 and N4 between which Clarke places it, and that if indeed Clarke's sequence N1-N4 is valid it must be based on subjective criteria, provenance or decoration. The general mixture of Clarke's N and S types among groups 1,2,4,5 and 8 above also questions the validity of the southern sequence. It is emphasized that this study has been based solely on the objective criteria of shape and size, and in some cases more subjective criteria, or data concerning provenance and decoration may be more important.

(c) The Eastern European projectile points (see Figure 1.21) considered form a single group at 43% similarity. The main groups seem to be defined by the 74% phenon, and are as follows:

1. A group of medium-sized, generally pear-shaped
projectile points DS1, SL1, JI1, JII1, O1
2. An elongated projectile point, FD1
3. A group of small, pointed projectile points
N10, NP1, SU1
4. A very broad projectile point, MD1.

1.10.10

POTVOL Pottery volumes (BASIC, designed for remote terminal use)

This program is a remote terminal version of the algorithm used in ARCJW005 to calculate the volume of wheel-made pots, and in ARCJW019 to calculate volume ratios during the comparison of two pots. The program is fully described in the chapter on Pottery Classification (see paragraph 4.4.5).

Statistics Bibliography

- Ammerman, A., 1972 The new archaeology
The Listener 87(2236), 143-144
- Ascher, M., 1959 A mathematical rationale for graphical
 seriation
American Antiquity 25(2), 212-214
- Ascher, M., 1966 S.20. Chronological ordering
C.Hum
 in *Directory of Scholars Active 1966/67*,
 234
- Ascher, M. and Ascher, R., 1963 Chronological ordering by computer
American Anthropologist 65(5), 1045-1052
- Belous, R.E., 1953 The central California chronological
 sequence re-examined
American Antiquity 18(4), 341-353
- Berry, B.J.L. and Marble, D.F. (eds), 1968
Spatial analysis
A reader in statistical geography
 Prentice-Hall, New Jersey 512 pp.
 (see also Chorley, R.J. and Haggett, P.,
 1968; Harbaugh, J.W. and Preston, F.W.,
 1968)
- Binford, L.R., 1964 A consideration of archaeological research
 design
American Antiquity 29(4), 425-441
- Binford, L.R. and Binford, S.R., 1966
 A preliminary analysis of functional
 variability in the Mousterian of Levallois
 facies
American Anthropologist 68(2), 238-295
 (Clark, J.D. and Howell, F.C. (eds) *Recent
 studies in palaeoanthropology*)
- Binford, S.R. and Binford, L.R. (eds), 1968
New perspectives in archaeology
 Aldine Publishing Co., Chicago 373 pp.
 (see also Cowgill, G.L., 1968c;
 Hill, J.N., 1968;
 Longacre, W.A., 1968;
 Whallon, R., 1968;
 Williams, B.J., 1968)

- Bohmers,A., 1963
A statistical analysis of flint artifacts
in Brothwell,D. and Higgs,E. (eds)
Science in archaeology 469-481
Thames and Hudson, London
- Bordes,F., 1964
Statistics and archaeology:
The use and interpretation of elementary
statistics in the Palaeolithic of France
Research seminar on statistics and archaeology
Institute of Archaeology, London, 30 May 1964
- Bordes,F. and de Sonneville-Bordes,D., 1970
The significance of variability in Palaeolithic
assemblages
World Archaeology 2(1), 61-73
- Borillo,M., 1972
CL1. Anthropologie, Archeologie, Histoire
de L'Art
C.Hum. 6(3), 179
in Humanities Programs Available
- Brainerd,G.W., 1951
The place of chronological ordering in
archaeological analysis
American Antiquity 16(4), 301-313
- Brothwell,D., 1969
Stones, pots and people: a plea for statistical
caution
in Brothwell,D. and Higgs,E. (eds), 1969
Science in archaeology 669-679
Thames and Hudson, London
- Brown,J.A. and Freeman,L.G., 1964
A UNIVAC analysis of sherd frequencies from
the Carter Ranch Pueblo, Eastern Arizona
American Antiquity 30(2), 162-167
- Chenhall,R.G., 1966
The logic of models used for processing
archaeological data on computers
*Proceedings of the international symposium
on mathematical and computational methods
in the social sciences*
International Computation Center, Rome
- Chenhall,R.G., 1968a
The impact of computers on archaeological
theory:
an appraisal and projection
C.Hum. 3(1), 15-24
- Chenhall,R.G., 1968b
Review of
Cowgill,G.L., 1967
Computer applications in archaeology
Computing Reviews 9, 187
- Chorley,R.J., 1964
Geography and analogue theory
Annals of the Association of American Geographers
54, 127-137

- Chorley, R.J. and Haggett, P., 1968
Trend-surface mapping in geographical research
in Berry, B.J.L. and Marble, D.F. (eds), 1968
Spatial analysis
A reader in statistical geography 195-217
Prentice-Hall, New Jersey
- Chouraqui, E., 1972
CL4/5. Archeologie, Histoire de l'Art
C.Hum. 6(3), 179-180
in Humanities Programs Available
- Clarke, D.L., 1963
Matrix analysis and archaeology
Nature 199(4895), 790-792
- Clarke, D.L., 1968
Analytical archaeology
Methuen, London 684 pp.
- Cloak, F.T.Jr., 1967
The computer in archaeology
Computer studies in the humanities and verbal behaviour 1(2), 61-69
(see also Cowgill, G.L., 1970;
Wilcock, J.D., 1969)
- Cochran, W.G., 1952
The χ^2 test of goodness of fit
Annals of Mathematical Statistics 23, 315-345
- Cole, A.J. and Wishart, D., 1970
An improved algorithm for the Jardine-Sibson method of generating overlapping clusters
Computer Journal 13(2), 156-163
- Collins, D., 1970
Stone artefact analysis and the recognition of culture traditions
World Archaeology 2(1), 17-27
- Cooley, J.W. and Tukey, J.W., 1965
An algorithm for the machine calculation of complex Fourier series
Math. Comp. 19, 297-301
- Cowgill, G.L., 1967a
Computer applications in archaeology
Fall Joint Computer Conference 1967
Proc. AFIPS 31, 331-337
Thompson Book Co., Washington
(see also Chenhall, R.G., 1968b)
- Cowgill, G.L., 1967b
Computer applications in archaeology
C.Hum. 2(1), 17-23
- Cowgill, G.L., 1968a
Archaeological applications of factor, cluster and proximity analysis
American Antiquity 33(3), 367-375
- Cowgill, G.L., 1968b
Review of
Hole, F. and Shaw, M., 1967
Computer analysis of chronological seriation
American Antiquity 33(4), 517-519

- Cowgill, G.L. 1968c
Computer analysis of archaeological data
Teotihuacan, Mexico
in Binford, S.R. and Binford, L.R. (eds), 1968
New perspectives in archeology 143-50
Aldine Publishing Co., Chicago
- Cowgill, G.L., 1970
Review of
Cloak, F.T. Jr, 1967
The Computer in archaeology
*Computer studies in the humanities and
verbal behaviour* 1(2), 61-69
C.Hum. 4(3), 221
- Craytor, W.B. and Johnson, L., 1968
Refinements in computerised item seriation
Bulletin 10
Museum of Natural History, University of
Oregon
- Deetz, J.D.F., 1965
*The dynamics of stylistic change in Arikara
ceramics*
Illinois Studies in Anthropology 4
University of Illinois Press, Urbana
- Deetz, J.D.F. and Dethlefsen, E., 1965
The Doppler effect and archaeology:
a consideration of the spatial aspects of
seriation
Southwestern Journal of Anthropology
21(3), 196-206
- Dempsey, P. and Baumhoff, M., 1963
The statistical use of artifact distributions
to establish chronological sequence
American Antiquity 28(4), 496-509
- Doran, J.E., 1967
Electronic computers and archaeology:
a computer scientist's viewpoint
Research Seminar on archaeology
Institute of Archaeology, London, 23 January
1967
- Doran, J.E., 1970a
Systems theory, computer simulations and
archaeology
World Archaeology 1(3), 289-298
- Doran, J.E., 1970b
Archaeological reasoning and machine reasoning
Archeologie et calculateurs 57-69
Centre National de la Recherche Scientifique
- Doran, J.E., 1971
Computer analysis of data from the La Tène
cemetery at Münsingen-Rain
in Hodson, F.R., Kendall, D.G. and Tautu, P.
(eds), 1971
*Mathematics in the archaeological and
historical sciences* 422-431
Edinburgh University Press, Edinburgh

- Doran, J.E. et al., 1972a The new archaeology
The Listener 87(2234), 68-70
- Doran, J.E., 1972b Correspondence in
Newsletter of Computer Archaeology 7(3),
 3-4
- Doran, J.E. and Hodson, F.R., 1966 A digital computer analysis of Palaeolithic
 flint assemblages
Nature 210(5037), 688-689
- Doran, J.E. and Hodson, F.R., 1971 An evaluation of statistical applications
Newsletter of Computer Archaeology 7(2),
 1-2
- Ellisseeff, V., 1965 Possibilités du scalogramme dans l'étude
 des bronzes Chinois archaïques
*Comptes-rendus du séminaire sur les modèles
 mathématiques dans les sciences sociales*
 1964-5
 Ecole Pratique des Hautes Etudes, Paris
Mathématiques et Sciences Humaines 11, 1-10
- Fernandez de la Vega, W., 1972 CL2/3. Archaeology
C.Hum. 6(3), 179
 in Humanities Programs Available
- Fisher, R.A., 1938 *Statistical methods for research workers*
 Oliver and Boyd, Edinburgh and London
 354 pp.
- Fruchter, B., 1954 *Introduction to factor analysis*
 D. van Nostrand, Princeton, N.J.
- Gardin, J.C., 1958 Four codes for the description of artifacts,
 an essay in archaeological technique and
 theory
American Anthropologist 60(2), 335-357
- Gardin, J.C., 1969 S93. Méthodes mathématiques et linguistiques
 en archéologie
C.Hum.
 in *Directory of Scholars Active 1968/9(2)*,
 141
- Gelfand, A.E., 1971a Seriation methods for archaeological
 materials
American Antiquity 36(3), 263-274
- Gelfand, A.E., 1971b Rapid seriation methods with archaeological
 applications
 in Hodson, F.R., Kendall, D.G. and Tautu, P. (eds),
 1971
*Mathematics in the archaeological and historical
 sciences* 186-201
 Edinburgh University Press, Edinburgh

- Gower, J.C., 1966
Some distance properties of latent root and vector methods used in multivariate analysis
Biometrika 53(3&4), 325-338
- Gower, J.C., 1971
Statistical methods of comparing different multivariate analyses of the same data in Hodson, F.R., Kendall, D.G. and Tautu, P. (eds), 1971
Mathematics in the archaeological and historical sciences 138-149
Edinburgh University Press, Edinburgh
- Graham, J.M., 1970 with a note by Roe, D.
Discrimination of British Lower and Middle Palaeolithic handaxe groups using canonical variates
World Archaeology 1(3), 321-342
- Harbaugh, J.W. and Preston, F.W., 1968
Fourier series analysis in geology in Berry, B.J.L. and Marble, D.F. (eds), 1968
Spatial analysis
A reader in statistical geography 218-238
Prentice-Hall, New Jersey
- Harman, H.H., 1967
Modern factor analysis
The University of Chicago Press, Chicago and London 474 pp.
- Hill, J.N., 1968
Broken K Pueblo: patterns of form and function in Binford, S.R. and Binford, L.R. (eds), 1968
New perspectives in archeology 103-142
Aldine Publishing Co., Chicago
- Hiorns, R.W., 1970
Review of Hope, K., 1969
Methods of multivariate analysis
Handbook of multivariate methods programmed in ATLAS Autocode
University of London Press 288 pp.
Computer Journal 13(1), 109
- Hiorns, R.W., 1971
Statistical studies in migration in Hodson, F.R., Kendall, D.G. and Tautu, P. (eds), 1971
Mathematics in the archaeological and historical sciences 291-302
Edinburgh University Press, Edinburgh
- Hodson, F.R., 1969a
Searching for structure within multivariate archaeological data
World Archaeology 1(1), 90-105
- Hodson, F.R., 1969b
Classification by computer in Brothwell, D. and Higgs, E. (eds), 1969
Science in archaeology 649-660
Thames and Hudson, London

- Hodson, F.R., 1970
Cluster analysis and archaeology: Some new developments and applications
World Archaeology 1(3), 299-320
- Hodson, F.R., 1971
Numerical typology and prehistoric archaeology
in Hodson, F.R., Kendall, D.G. and Tautu, P. (eds), 1971
Mathematics in the archaeological and historical sciences 30-45
Edinburgh University Press, Edinburgh
- Hodson, F.R., Kendall, D.G. and Tautu, P. (eds), 1971
Mathematics in the archaeological and historical sciences
Edinburgh University Press, Edinburgh 565 pp.
(see also Doran, J.E., 1971;
Gelfand, A.E., 1971b;
Gower, J.C., 1971;
Hiorns, R.W., 1971;
Hodson, F.R., 1971;
Kendall, D.G., 1971;
Kruskal, J.B., 1971;
Moberg, C.A., 1971;
Moore, R.E.M., 1971;
Orton, C.R., 1971;
Sibson, R., 1971b, 1971c;
Spaulding, A.C., 1971;
Wilcock, J.D., 1971;
Wilkinson, E.M., 1971)
- Hodson, F.R., Sneath, P.H.A. and Doran, J.E., 1966
Some experiments in the numerical analysis of archaeological data
Biometrika 53(3&4), 311-324
- Hoel, P., 1962
Introduction to mathematical statistics
Wiley and Sons
- Hole, F., 1966
Chronological ordering by computer
MASCA Newsletter 2(1), 2
- Hole, F. and Shaw, M., 1967a
Computer analysis of chronological seriation
Rice University Studies Monographs 53(3)
Houston 166 pp.
- Hole, F. and Shaw, M., 1967b
S40. Computer analysis of chronological change in archeological data
C.Hum.
in *Directory of Scholars Active 1967/68*(1), 90
- Hymes, D. (ed.), 1965
The use of computers in anthropology
Mouton & Co., The Hague, London and Paris 558 pp.
(see also Needham, R.M., 1965)

- Ihm, P., 1961
Classification automatique des objets
de l'Âge du Bronze
*Comptes-rendus du séminaire sur les modèles
mathématiques dans les sciences sociales*,
1960-1 28-33
Ecole Pratique des Hautes Etudes, Paris
- Jardine, C.J., Jardine, N. and Sibson, R., 1967
The structure and construction of taxonomic
hierarchies
Mathematical Biosciences 1, 173-179
- Jardine, N., 1970
Algorithms, methods and models in the
simplification of complex data
Computer Journal 13(1), 116-117
- Jardine, N. and Sibson, R., 1968
The construction of hierarchic and non-
hierarchic classifications
Computer Journal 11(2), 177-184
- Jardine, N. and Sibson, R., 1971
Choice of methods for automatic classification
Computer Journal 14(4), 404-406
- Kamenetskij, I.S., 1965
Datation of layers from percentages of
pottery types
in *Archaeology and the natural sciences* 302-307
Nauka, Moscow
- Kendall, D.G., 1963
A statistical approach to Flinders Petrie's
sequence dating
*Bulletin of the International Statistics
Institute* 40, 657-680
34th session, Ottawa
- Kendall, D.G., 1969
Some problems and methods in Statistical
archaeology
World Archaeology 1(1), 68-76
- Kendall, D.G., 1970a
Operation Speckled band
Science and Archaeology 2/3, 8 N2/6
- Kendall, D.G., 1970b
A mathematical approach to seriation
Phil. Trans. Roy. Soc. Lond. A.269, 125-135
- Kendall, D.G., 1971
Seriation from abundance matrices
in Hodson, F.R., Kendall, D.G., and Tautu, P.
(eds), 1971
*Mathematics in the archaeological and
historical sciences* 215-252
Edinburgh University Press, Edinburgh
- Kerrich, J.E. and Clarke, D.L., 1967
Notes on the possible misuse and errors
of cumulative percentage frequency graphs
for the comparison of prehistoric artefact
assemblages
Proc. Prehist. Soc. N.S.33, 57-69

- Kovalevskaja, V.B., 1965 The use of statistical methods for the study of large archaeological collections in *Archaeology and the natural sciences* 286-300
Nauka, Moscow
- Krumbein, W.C., 1959 Trend Surface analysis of contour-type maps with irregular control-point spacing *Journal of Geophysical Research* 64(7), 823-834
- Kruskal, J.B., 1964a Multidimensional scaling by optimizing goodness-of-fit to a nonmetric hypothesis *Psychometrika* 29(1), 1-27
- Kruskal, J.B., 1964b Nonmetric multidimensional scaling: a numerical method *Psychometrika* 29(2), 115-129
- Kruskal, J.B., 1971 Multi-dimensional scaling in archaeology: time is not the only dimension in Hodson, F.R., Kendall, D.G. and Tautu, P. (eds), 1971 *Mathematics in the archaeological and historical sciences* 119-132
Edinburgh University Press, Edinburgh
- Kuzara, R.S., Mead, G.R. and Dixon, K.A., 1966 Seriation of anthropological data: a computer program for matrix ordering *American Anthropologist* 68(6), 1442-1455
- Lance, G.N. and Williams, W.T., 1965 Computer programs for monothetic classification *Computer Journal* 8(3), 246-249
- Lance, G.N. and Williams, W.T., 1966 Computer programs for hierarchic polythetic classification ("similarity analyses") *Computer Journal* 9(1), 60-64
- Lance, G.N. and Williams, W.T., 1967a A general theory of classificatory sorting strategies
I. Hierarchical systems *Computer Journal* 9(4), 373-380
- Lance, G.N. and Williams, W.T., 1967b Note on the classification of multi-level data *Computer Journal* 9(4), 381-382
- Lance, G.N. and Williams, W.T., 1967c A general theory of classificatory sorting strategies
II. Clustering systems *Computer Journal* 10(3), 271-277

- Lance, G.N. and Williams, W.T., 1968
Note on a new information-statistic
classificatory program
Computer Journal 11(2), 195
- Lance, G.N. and Williams, W.T., 1971
A note on a new divisive classificatory
program for mixed data
Computer Journal 14(2), 154-155
- Lipe, W.D., 1964
Comment on Dempsey and Baumhoff's "The
statistical use of artifact distributions
to establish chronological sequence"
American Antiquity 30(1), 103-104
(see Dempsey, P. and Baumhoff, M., 1963)
- Longacre, W.A., 1968
Some aspects of prehistoric society in
East-Central Arizona
in Binford, S.R. and Binford, L.R. (eds),
1968
New perspectives in archeology 89-102
Aldine Publishing Co., Chicago
- McBurney, C.B.M., 1964
Statistics and archaeology: Some observed
examples of multi-variate change with
time
*Research seminar on statistics and
archaeology*
Institute of Archaeology, London, 30 May
1964
- McPherron, A., 1963
Programming the IBM 7090 for optimizing
taxonomy in archaeology
*62nd annual meeting of the American
Anthropological Association, Chicago*
- Mahalanobis, P.C., 1936
On the generalised distance in statistics
Proc. Nat. Inst. Sci. India 2, 49
- Martin, P.S., 1966
S6. Evolution of social organization in
Prehistoric Arizona
C. Hum.
in *Directory of Scholars Active 1966/67*,
230
- Massey, F.J. Jr., 1952
Distribution table for the deviation
between two simple cumulatives
Annals of Mathematical Statistics 23,
435-441
- Matthews, J., 1963
Application of matrix analysis to
archaeological problems
Nature 198, 930-934
Matrix analysis and archaeology - a reply
Nature 199, 790-792
- Meighan, C.W., 1959
A new method for the seriation of
archaeological collections
American Antiquity 25(2), 203-211

- Mellars, P.A., 1970 Some comments on the notion of 'functional variability' in stone-tool assemblages
World Archaeology 2(1), 74-89
- Millon, R., 1966 S28. Urbanization at Teotihuacan, Mexico
C.Hum.
in Directory of Scholars Active 1966/67, 237
- Millon, R. and Cowgill, G.L., 1968 S28. Urbanization at Teotihuacan, Mexico
C.Hum.
in Directory of Scholars Active 1968/69(1), 314
- Moberg, C.A., 1961 Mangder au fornfynd, kring aktuella tendenser I
Arkeologisk Metodik, Acta Universitatis G  ttingensis 67(1), 1-31
- Moberg, C.A., 1966 An experiment with determination of technological traits of assemblages of flint blades and flakes
Proceedings of the international symposium on mathematical and computational methods in the social sciences
International Computation Center, Rome
- Moberg, C.A., 1971 Closing address: archaeological context and mathematical methods
in Hodson, F.R., Kendall, D.G. and Tautu, P. (eds), 1971
Mathematics in the archaeological and historical sciences
551-562
Edinburgh University Press, Edinburgh
- Needham, R.M., 1965 Computer methods for classification and grouping
in Hymes, D. (ed.), 1965
The use of computers in anthropology 345-356
Mouton & Co., The Hague, London and Paris
- Petrie, W.M. Flinders, 1899 Sequences in Prehistoric remains
J. Roy. Anthropol. Inst. 29, 295-301
- Rao, C.R., 1966 Discriminant function between composite hypotheses and related problems
Biometrika 53(3&4), 339-345
- Renfrew, C. and Sterud, G., 1969 Close-proximity analysis: a rapid method for the ordering of archaeological materials
American Antiquity 34(3), 265-277
- Rhoads, J.G., 1968 S85. Systems for analytical archaeology
C.Hum.
in Directory of Scholars Active 1968/69(1), 319

- van Rijsbergen, C.J., 1970 Algorithm 47: A clustering algorithm
Computer Journal 13(1), 113-115
(designed in collaboration with Jardine,
N. and Jardine, C.J.)
- Robinson, W.S., 1951 A method for chronologically ordering
archaeological deposits
American Antiquity 16(4), 293-301
- Robinson, W.S. and Brainerd, G.W., 1952/53
Robinson's coefficient of agreement -
a rejoinder
American Antiquity 18(1), 60-61
- Roe, D.A., 1964a Statistics and archaeology: metrical and
statistical analysis of handaxe groups in
the British Lower Palaeolithic
*Research Seminar on statistics and
archaeology*
Institute of Archaeology, London, 30 May 1964
- Roe, D.A., 1964b The British Lower and Middle Palaeolithic:
some problems, methods of study and
preliminary results
Proc. Prehist. Soc. N.S.30, 245-267
- Roe, D.A., 1968 British Lower and Middle Palaeolithic
handaxe groups
Proc. Prehist. Soc. N.S.34, 1-82
- Rowlett, R.M., 1966 S17. Proportion analysis of Hallstatt and
La Tène spearheads
C. Hum.
in *Directory of Scholars Active 1966/67*, 233
- Rowlett, R.M. and Robbins, M.C., 1969
S89. The selection of representative sites
in archaeology
C. Hum.
in *Directory of Scholars Active, 1968-69(2)*,
139
- Sackett, J.R., 1966 Quantitative analysis of Upper Palaeolithic
stone tools
American Anthropologist 68(2), 356-394
- Shepard, R.N., 1962 The analysis of proximities: multidimensional
scaling with an unknown distance function
Psychometrika 27(2), 125-140 and 219-246
- Sibson, R., 1971a Some observations on a paper by Lance and
Williams
Computer Journal 14(2), 156-157
- Sibson, R., 1971b Computational methods in cluster analysis
in Hodson, F.R., Kendall, D.G. and Tautu, P.,
(eds), 1971
*Mathematics in the archaeological and
historical sciences* 59-61
Edinburgh University Press, Edinburgh

- Sibson, R., 1971c
Some thoughts on sequencing methods
in Hodson, F.R., Kendall, D.G. and Tautu, P.
(eds), 1971
*Mathematics in the archaeological and
historical sciences* 263-266
Edinburgh University Press, Edinburgh
- Smirnov, N., 1948
Tables for estimating the goodness of fit
of empirical distributions
Annals of Mathematical Statistics 19,
279-281
- Sneath, P.H.A., 1967
Trend surface analysis of transformation
grids
Journal of Zoology 151, 65-122
- Sneath, P.H.A. and Sokal, R.R., 1962
Numerical taxonomy
Nature 193(4818), 855-860
- Sokal, R.R. and Sneath, P.H.A., 1963
Principles of numerical taxonomy
Freeman, San Francisco and London 359 pp.
- Spaulding, A.C., 1953
Statistical techniques for the discovery of
artifact types
American Antiquity 18(4), 305-313
- Spaulding, A.C., 1960
Statistical description and comparison of
artifact assemblages
in Heizer, R.F. and Cook, S.F. (eds), 1960
*The application of quantitative methods in
archaeology*
60-92
Viking Fund Publications in Archaeology 28
Quadrangle Books, Chicago
- Spaulding, A.C., 1971
Some elements of quantitative archaeology
in Hodson, F.R., Kendall, D.G. and Tautu, P.
(eds), 1971
*Mathematics in the archaeological and
historical sciences* 3-16
Edinburgh University Press, Edinburgh
- Thomas, D.H., 1969
S92. Applications of numerical taxonomy to
archaeology
C. Hum.
in Directory of Scholars Active 1968/69(2),
140
- Thomas, D.H., 1971
On the use of cumulative curves and
numerical taxonomy
American Antiquity 36(2), 206-209
- Tugby, D.J., 1965
Archaeological objectives and statistical
methods: a frontier in archaeology
American Antiquity 31(1), 1-16

- Tugby,D.J., 1969
Archaeology and statistics
in Brothwell,D. and Higgs,E. (eds), 1969
Science in archaeology 635-648
Thames and Hudson, London
- Whallon,R., 1968
Investigations of Late Prehistoric social
organisation in New York state
in Binford,S.R. and Binford,L.R. (eds),
1968
New perspectives in archeology 223-244
Aldine Publishing Co., Chicago
- Wilcock,J.D., 1969
Review of
Cloak,F.T.Jr, 1967
The computer in archaeology
in *Computer studies in the humanities and
verbal behaviour* 1(2), 61-69
Computing Reviews 10, 370-371 (ref.17,207)
- Wilcock,J.D., 1970
Review of
Clarke,D.L., 1968
Analytical archaeology
Methuen, London 684 pp.
Science and Archaeology 1, 29
- Wilcock,J.D., 1970
Our man in Mamaia
Science and Archaeology 2/3, 12-16
- Wilkinson,E.M., 1971
Archaeological seriation and the travelling
Salesman problem
in Hodson,F.R., Kendall,D.G. and Tautu,P.
(eds), 1971
*Mathematics in the archaeological and
historical sciences* 276-283
Edinburgh University Press, Edinburgh
- Williams,B.J., 1968
Establishing cultural heterogeneities in
settlement patterns: an ethnographic example
in Binford,S.R. and Binford,L.R. (eds), 1968
New perspectives in archeology 161-170
Aldine Publishing Co., Chicago
- Williams,W.T., Clifford,H.T. and Lance,G.N., 1971
Group-size dependence: a rationale for
choice between numerical classifications
Computer Journal 14(2), 157-162
- Williams,W.T., Lance,G.N., Dale,M.B. and Clifford,H.T., 1971
Controversy concerning the criteria for
taxonomic strategies
Computer Journal 14(2), 162-165

Chapter 2

THE USE OF INFORMATION RETRIEVAL IN ARCHAEOLOGY

2.1 *Information retrieval* is employed today in many fields. It was inevitable that the computer should be applied to the problem of information retrieval in archaeology. Every excavation produces a large amount of information, and yet usually only a small "representative" sample of this ever gets published. The remainder is often lost or at best bequeathed to a museum, where it may (but only in the more efficient museums) be catalogued on a card file. The first information application in archaeology is the excavation record. In the traditional situation publication is often delayed for many years after the close of the excavation. The more progressive archaeologists are now beginning to use remote terminals on site for the routine recording of excavated material (a suite of remote terminal programs has been developed for this thesis, in collaboration with Mr P. Buckland of Doncaster Museum and the University of Birmingham, for use on the DANVM Roman fort site in Doncaster; and with the Department of the Environment Archaeology Section, for use on the Houses of Commons Extension site, London). The accumulated records are invaluable for information retrieval runs and straight listings, and speed the publication of the material.

The second information retrieval application in archaeology is the museum catalogue. The more progressive museum curators are beginning to automate their catalogues. For this study a museum was selected which had never compiled a catalogue of any sort, not even an accessions list, but which had a progressive curator (Ribchester museum, Lancashire, curator Mrs N.M. Dixon). The entire Roman collection from the BREMETENNACVM fort site has been reorganised by Mrs Dixon, using modern display techniques, and a large number of models have been constructed as visual aids. As a result, the number of visitors has soared and the educational role of the museum is becoming important, with an

increase in number of school parties. Under these circumstances it was decided that a published catalogue giving all items on display in the museum was desirable. The material includes Roman inscriptions, coins, pottery, glass and metalwork and models. All items will be entered into a computer information retrieval scheme as they become available (inscriptions and models have so far been processed). The data will be sorted into various orders (e.g. alphabetical order by material, bibliographic reference, accession order, or position in the museum) and printed by line printer for offset-litho reproduction. As new items are acquired by the museum the computer file is amended, and a new edition of the catalogue can at any time be run-off.

The third information retrieval application in archaeology is the large body of specialist information. Under this category the following have been studied:

(a) The Roman Inscriptions of Britain. The data for this has been taken from Collingwood and Wright, 1965, and the magnetic tape record constitutes in law a single copy for research purposes only. Copyright law forbids direct publication of the material, although new interpretations of the data may be published. Special permission from the publishers and author has been obtained for direct publication of the Ribchester inscriptions in the museum catalogue. The work is described in section 2.4 and paragraphs 2.6.2 and 2.6.6.

(b) Cave finds from the British Isles (as Hon. Recorder for Archaeology and Palaeontology of the Cave Research Group of Great Britain).

(c) British petroglyphs, in association with Dr M.J.Walker of the University of Edinburgh. The work is described in paragraphs 2.3.28, 2.6.2 and 2.6.6.

(d) Clay pipe varieties, which are useful for dating purposes, in

association with Dr F.Celoria of the University of Keele. Coding is described in paragraphs 2.6.2 and 2.6.6.

(e) Masons' marks, in association with Dr F.Celoria of the University of Keele. For full description see paragraph 1.10.3

(f) Pottery types; a typical record consists of alphanumeric information, plus graphic information for drawing the pot, the volume of the pot, scale factor and profile code for the calculation of similarity coefficients. For full description see paragraph 4.4.3).

(g) Eastern European projectile points, in association with P.Allsworth-Jones of Emmanuel College, Cambridge. A typical record consists of alphanumeric information, plus graphic information for drawing the projectile point, the area of the plan view of the projectile point scale factor and Fourier spectrum for the calculation of similarity coefficients. For full description see paragraphs 1.10.5 and 1.10.6.

2.2

The applications will be discussed in this chapter under the following sections:

2.3 General

2.4 Roman Inscriptions

2.5 Ribchester Museum Catalogue

2.6 Information retrieval programs developed

2.3

General information retrieval applications

Under this heading will be discussed information retrieval in archaeology and the museum, as well as some specialist topics not discussed in sections 2.4 and 2.5.

2.3.1

Computer information retrieval was the inheritor of numerous card-handling techniques, the developments of the original Hollerith card system invented in 1890. Jonker (1963) describes the Termatrix

system which uses a large card with 10,000 possible hole positions. The method has been developed by the Anson-Jonker company as their Geo-Stat technique. The method uses plastic cards in which holes are drilled using a precision drilling machine. The PEEKABOO or *optical coincidence* method is used to determine which holes are common in a series of cards. Either a *normal* or an *inverted* file system may be in use. In a normal file each card represents one item and each hole a specific attribute; this corresponds to the card files kept in some museums, which have a card describing each item. Thus items which have a given attribute may be found by withdrawing cards from the pack which have a hole in the corresponding position.

Edge-punched cards are often used in this sort of application, the cards being withdrawn from the pack by inserting a needle through the whole pack at a particular hole, when the cards punched in this position drop out. In an inverted file each card represents an attribute and each hole a specific item. The item is defined in terms of attributes by selecting the required attribute cards, stacking them together in the drilling machine, and drilling them all together at the point defining the item. To find all items which have a given combination of attributes, the required attribute cards are withdrawn from the pack and the items which have all these attributes are determined by optical coincidence.

The Anson-Jonker equipment has a reading unit with cursor, which may be used to select any required hole out of the 10,000. There is also a series-scanning unit which may be used to scan a whole card, and a photocell-activated counter which records the number of optical coincidence holes found. The full 10,000 holes may be scanned in 8 seconds. Transparent map overlays may be used, and holes drilled to indicate archaeological find sites; the hole matrix may be made to

coincide with the National Grid, and thus the sensor can be made to read off the NGRs of all finds automatically; this is a very useful feature for the archaeologist, and an additional mode of working to the normal and inverted file systems described above.

2.3.2

Atkinson and Oswald describe some London clay tobacco pipes, giving a chronology of bowl types. The various types are illustrated, and these diagrams have been used in the study of clay pipe typology in this thesis. The profiles are recorded by d-Mac pencil follower, then rotated and scaled to standard orientation and height with adjustment of scale factor. A typical magnetic tape record consists of alphanumeric data read from cards; scale factor; and rotated, smoothed and scaled coordinates for drawing the clay pipe. The system is fully described in paragraphs 2.6.2 and 2.6.6.

2.3.3

Bordaz and Bordaz (1965/6) describe the use of the Termatrex system (see paragraph 2.3.1 above). They use an inverted file of Termatrex cards, converting the data to standard Hollerith punched cards or to paper tape for computer input, and computer output on cards or paper tape to Termatrex cards.

2.3.4

The first use of electronic data processing, rather than machine-sorted cards, in information retrieval for archaeology was probably by Gardin in 1965. He used the computer to sort and retrieve 10,000 documents concerning Assyrian-Cappadocian trade records, giving relations between individuals, the purpose of the relations (barter, trade, loans, transport, etc.), the products concerned, and the mention of place names as the temporary locations of given merchants. Patronymics are not often given, and the problem of homonyms is severe. The computer is used to infer the locations of co-addressees or co-senders from

the known locations of one of their number. An algorithm is given which attempts to deduce the probable location of merchants from direct and indirect evidence, e.g. the invariable direction of tin deliveries. The flowchart given, however, has many illogicalities, and could not be implemented on the computer as shown. Gardin in fact states that his flowchart is "problem-oriented" rather than "machine-oriented". In the discussion of this paper the remark is made that "the computer is beginning to be important in archaeology".

2.3.5

One of the major workers in museum information retrieval in Britain is Geoffrey Lewis, curator of Sheffield City Museum, and Chairman of the Information Retrieval Group of the Museums Association (IRGMA). In a 1965 paper he gives a general description of information retrieval methods (Hollerith, edge-punched and PEEKABOO cards), and presents some thoughts on a national museum index. Suggested headings for museum classifications are subject (e.g. archaeology); name of specimen; material; find-spot (including grid reference); date of manufacture (absolute and relative dating); maker and place of manufacture; sex (for biological specimens); historical associations; and other information. A national lead is required in the organising of a standard index.

2.3.6

Green (1966) reports FORTRAN IV software written for the IBM 7040 machine fitted with magnetic tape units, disc files and a remote teletype in the museum laboratory. The object is to develop a catalogue-research system for the control of the collection of the Anthropology Museum of the University of Missouri, with future remote linkage to other museum computer systems. All items are to be recorded on magnetic tape. Statistical and data-handling programs useful

in anthropological research, with special reference to large archaeological collections, are to be set up.

2.3.7

Gundlach (1966) describes 1966 ARDOC, a system of information storage and retrieval in archaeology. The system has been used to store pottery data as specified by Ankel (1966/67, see chapter on Pottery Classification, paragraph 4.3.8).

2.3.8

Chenhall (1966,1967) reports FORTRAN and machine code software for the CDC 3400 and IBM 1440 machines for museum cataloguing. The objective is to provide lists of ethnological and archaeological museum collections organised by type of items, by culture, by location, by material, and by method of manufacture. Plain-language description, recording and listing, which are most important from the archaeologist's point of view, are discussed right at the end of the 1967 paper, but the actual implementation uses a complex numerical code. Chenhall states that the requirements for the storage, retrieval and analysis of archaeological data are that the artefacts should be described within a series of models which are:

- (a) Sufficiently oriented to the "real world" to be useful;
- (b) open-ended so as to be useful in situations as yet undefined;
- and (c) specific enough to allow conversion into machine language.

2.3.9

A series of papers edited by Lewis are published in the Museums Journal for 1967. They include a report of the 1967 colloquium on information retrieval at the Sheffield City Museum; Miss C.Lavell on "The Council for British Archaeology and Information Retrieval", Dr F.Perring (Monks Wood Experimental Station) on "The requirements of the research worker in biology", Colin Renfrew on "The requirements of the research worker in archaeology", H.Hislop on "Information retrieval and computer-printed indexes", J.Chaldecott on "The automatic type-

writer for computer input", J.Brailsford on "Indexing techniques in the Department of British and Medieval Antiquities, British Museum", A.Gailey on "Proposals for indexing folk material" and Y.Oddon (UNESCO - ICOM Documentation Centre) in "Information retrieval and ICOM". The papers by Lavell, Renfrew and Perring will be covered in paragraphs 2.3.10, 2.3.11 and 2.3.12 below. Hislop comments on the sale of printed indexes organised in sequence of reference or accession number, of location of items within the museum, of general classification, or of specialised classification. The Romano-British collection at the British Museum has been transcribed to 80-column cards from the original registers. This thesis describes the Ribchester Museum project (section 2.5) to list the catalogue by computer for publication and sale. Brailsford describes indexing techniques used in the Department of British and Medieval Antiques, British Museum. The present acquisition registers contain a detailed objective description, a small sketch, particulars of acquisition and archaeological information for each item. 80-column cards have been punched with this information. Oddon describes the use of the Filmorex microfilm retrieval system for both photographs and text. Microfilm systems have a lot to offer the archaeologist in the recording of documents and photographs of objects.

2.3.10

Miss C.Lavell (1967) mentions the archaeological bibliography published for many years by the Council for British Archaeology. An abstracting service is operated which, however, is selective, publishing only those articles which are judged to add something to the general archaeological picture. Routine excavations and articles of local interest only are omitted. A card index has been considered but not yet implemented. Work on an agreed terminology in archaeology had been commenced by CBA, but had to be terminated because some definitions

could not be agreed upon in the fluid state of archaeology.

2.3.11

Renfrew (1967) comments on the need for information on the provenance of an object, and its present location in the museum, and also on the need for a national museum index. Research workers in a particular topic need to be able to determine the museums containing items of interest to them. The computer is useful in the plotting of distribution maps (see paragraphs 2.6.3, 2.6.4 and 2.6.7).

Renfrew proposes different sections of a catalogue for biology, geology, archaeology, history, art, etc. The existence of marginal categories belonging to two or more sections is inconvenient but inevitable. The special requirements of the archaeologist include associations, material, authorities for identification, grid reference of the find, type and culture. Renfrew makes the point that archaeological codes of description can never have the finality of, for example, species names in biology. This raises the problem of archaeological "type", its advantages and disadvantages, and whether or not the concept of type could be abolished.

2.3.12

Perring describes the requirements of the research worker in biology in two articles (Perring, 1967; Soper and Perring, 1967). Biology is linked with archaeology via cave finds, which have been investigated from an information retrieval point of view in this study. Perring is of the opinion that electronic data processing should be a by-product of the initial requirement to label all items accurately. In this he criticises Sokal and Sneath (1966) in their assertion that the EDP system is the more important, but labels could be provided if desired. Perring uses 40-column punched cards, and stores taxonomic keys in the computer, using multi-access computing facilities to control a graph plotter. A card sorter and a tabulator are also in use. For paper tape input a Flexowriter (tape-controlled

teletype) controlled by a master tape is used, and tape-to-card conversion may be performed if desired. A specialised piece of equipment in use is a pattern-select sorter, which will sort on 8 columns of a card simultaneously. Perring deplores the lack of a catalogue in many museums. For biological collections in particular, any published catalogue becomes rapidly out of date, but it ought to be possible for a museum to supply on demand at a moderate price copies of part of the current computer file. Without a catalogue the curator cannot compile lists of omissions, and the collection becomes arbitrary.

2.3.13

Rensberger and Berry (1967) describe an automated system for the retrieval of museum data, using punched cards and magnetic tape. Stylised codes, which have disadvantages for the archaeologist, rather than plain-language codes, are used. The system allows serial and random access to normal and inverted files. The PEEKABOO method of matching records is used. The following files have been constructed:

- (a) bibliographic file (year of publication, publication, author)
- (b) specimen file (collection data: year of collection, collector;
definition data: description, classification)
- (c) locality file (location name, grid reference).

Each of the files has cross-reference keys to the other files. There are facilities for sorting and merging files, updating indexing and printing. A special synonym file checks for the use of synonyms, and finds the preferred nomenclature during information retrieval runs. Economics may dictate that a single punched card be used per item.

2.3.14

Chenhall (1968) analyses the features of museum information retrieval systems, making a critical comparison.

2.3.15

Hutton (1968) describes the incorporation of the PEEKABOO method

into an IBM 7090 program. The algorithm creates an 18-byte mask of the item's meaningful words using the INCLUSIVE OR logical operation. The mask is stored with the item. At search time, a question mask is constructed similarly from keywords and compared with each item mask using the AND logical operation. If the result does not equal the question mask then it means that one or more question keywords are not present in the item mask, and the item is rejected. However, an equality of the question mask and the AND result is inconclusive; the actual keywords of the item must be unpacked and compared with the question keywords individually.

2.3.16

Morrison (1968) reports FORTRAN software for the CDC 3600 machine equipped with disc storage which gives facilities for the storing, retrieving and indexing of information in a flexible manner. The information is recorded in alphanumeric, with minimum restriction on the format of text and keys. The system does not require the rearrangement of text or index as new material is added. As has been mentioned above the flexible handling of unrestricted alphanumeric strings is of prime importance in archaeology.

2.3.17

O'Connor, Voigt and Wyuker (1968) report a program to perform a cross-indexing function. The program handles 27 major categories of information, including the object's provenance, present location, museum accession number, measurements, date and material and other descriptive categories. Within each major category up to 160 minor descriptive terms may be specified, with the aim that the program will be adaptable to most archaeological sites. A logical search algorithm is available for the combination of descriptions, although it is not made clear whether this just involves the AND operation, or other operations as well.

Conventional archaeological terms are used, allowing cards to be punched direct from field records without intermediate coding, and making interpretation easy. The program is written in machine code for the IBM 360/65 equipped with 2311 disc pack.

2.3.18

Another major British worker in the field of museum information retrieval is C.H.Roads of the Imperial War Museum. In a 1968 paper he describes the use of 80-column punched cards and *aperture cards*, which are 80-column cards with a space for insertion of a microfilm photograph of a document or object. Aperture cards may be filed as normal punched cards and may be punched in those columns not taken up by the microfilm aperture. The future transcription of the cards onto magnetic discs for computer processing is discussed.

2.3.19

Rusco (1968) describes the work of the Nevada Archaeological Survey since 1966 in the setting up of a computer storage and retrieval system. Numeric codes are in use. The punched cards are processed by an SDS Sigma VII machine and sorted by each descriptor category for a complete printout, for use as an index. A card sorter is to be used to derive statistics. The code system allows 56 categories of data for each site, including site number, location, ecological and geographical information, site characteristics, the status of the archaeological investigation, and the presence or absence of types of artefacts. The system has been designed to be expandable.

2.3.20

Schneider (1968) reports FORTRAN IV software for the IBM 7040 machine equipped with remote terminal in the museum. The purpose is to provide a catalogue-research system for the archaeological and ethnological collections of the Museum of Anthropology, University of Missouri. Items are coded according to culture, present location, donor etc.,

and an index is produced by sorting the data using system software.

2.3.21

A major worker in U.S.A. in the field of museum information retrieval is Everett Ellin, Director of the Museum Computer Network, a consortium of 25 east-coast museums. The initial work is reported in five 1968 references and one 1969 reference. The software is written in PL/I for the IBM 360 range of machines equipped with disc unit and an IBM 2321 data cell. A preliminary systems design has been completed and a model data bank constructed from the records of several thousand art objects in museum collections in the New York area. The envisaged network will disseminate information including textual and visual records over a communication system for museum, education and research purposes. In his international survey of museum computer activity (1968c) Ellin mentions IRGMA, the Imperial War Museum, and the Sedgwick Museum of Geology (Cambridge). There is a good survey of the information retrieval field and associated (mainly statistical) projects, with a select bibliography. An appendix gives the names and addresses of international workers including Smith (Akhenaten temple project, see paragraph 6.6.1), Cutbill (Sedgwick Museum of Geology, Cambridge), Lewis (Sheffield City Museum), Perring (Monks Wood Experimental Station), Roads (Imperial War Museum), Gardin, Ankel (pottery classification, see paragraph 4.3.8), Gundlach, Benfer, Burton (graphics for museums, see paragraph 3.3.17), Chenhall, Dauterman (Sèvres porcelain, see paragraph 4.3.10), Ellin, Gaines, Green, Longacre and Schneider.

2.3.22

Newman (1969), in correspondence with Chenhall, describes what is probably the first use of a remote terminal on an archaeological site. Newman has been experimenting for several years with the direct recording of archaeological data in the field, in a form suitable for

computer input. The object is to use the computer as an archaeological tool while site excavations are in progress. The 1969 test linked a site in Hawaii via a communications satellite to a computer in Palo Alto, California, with successful results. Complete analytical summaries of the day's work were received the same evening. Although this is probably the first use of data transmission between a site and computer, it is not the first archaeological data transmission exercise. An earlier exercise used the British TELEX system to transmit computer-processed geophysical survey results in 1968 (Wilcock, 1968, South Cadbury Castle, see paragraph 5.2.22).

2.3.23

The IRGMA Newsletter was launched in 1969, under a steering committee consisting of the following:

Barker (British Museum Research Laboratory)

Compton (Tate Gallery)

Cutbill (Sedgwick Museum of Geology, Cambridge)

Franks (Manchester Museum)

Hallan and Roads (Imperial War Museum)

Stansfield (Dept of Museum Studies, Leicester)

Vickery (ASLIB)

Leamy (Office for Scientific and Technical Information)

Miss B.Capstick (Secretary, Museums Association) under the chairmanship of Lewis (Sheffield City Museum). IRGMA published draft proposals for an interdisciplinary museum cataloguing system in April 1969 (Lewis et al., 1969). This is a full list of codes used in the IRGMA pilot scheme, but it is not very suitable for the average archaeological site record, nor for the routine cataloguing of county and parish finds. The system is described by Lewis (1970/71). He also comments on the reluctance of some curators to employ computer methods, stating that failure to appreciate the significance of the

computer in a profession which is concerned with the collection, processing and dissemination of information can only diminish the status of the museum as an information source. There is evidence that the cost of preparing computer input documents is comparable to the cost of traditional methods for museum indexing. IRGMA was formed in 1967, and in 1969 came to the conclusion that a cataloguing system common to all museums in the U.K. would be unrealistic at the present time. Systems designed to meet local requirements and machines are more efficient, and a national cataloguing standard designed to accommodate all possible local requirements would be most unwieldy and uneconomic. Museum information often originates from a variety of sources outside the control of the museum, and of variable reliability and completeness. New hardware developments are likely to make present methods of storage obsolete. For compatibility the hardware, data structure and media must be compatible, and this is not feasible at present because different computer manufacturers have different standards (attempts have been made, with varying degrees of success, to standardise computers via such bodies as the International Standards Organization (ISO) and the European Computer Manufacturers' Association (ECMA), but one computer manufacturer, IBM, is so large compared with all the others that it can be, and is, a law unto itself). The IRGMA proposals were designed to be a common language for the communication of data between local systems. This is the only feasible way of obtaining transfer of data between systems at present. Just as high-level computer languages are accepted by all computers, the common data language can be translated into the local data format and vice versa using a translator specially designed for the local system. If all local systems have these individual translators then transfer of data between systems can be effected. Efficient and economic national

cataloguing, however, can only come when manufacturers standardise computer equipment.

2.3.24

Bergengren (1970/71) remarks on the shortcomings of card files. In Sweden paper tape has been used instead, using standard control codes on a master control tape. The records are analysed by CORSAIR II (Computer Oriented Reference System for Automatic Information Retrieval), and there is a possibility of multi-access being available in due course.

2.3.25

Cuisenier (1970/71) describes a hierarchical keyword system for information retrieval for use at the Musée des Arts et Traditions Populaires, Paris. A hierarchical keyword system has been used in the remote terminal information retrieval programs developed for this thesis (see paragraphs 2.6.13 and 2.6.17).

2.3.26

Chenhall (1970) reports PL/I software for the IBM 360/50 which is to be used for the Arkansas Archaeological Data Bank. The Arkansas Archaeological Survey has professional archaeologists in all eight constituent colleges of the University of Arkansas. The purpose of the project is to store, retrieve and analyse the data produced by these and other archaeologists. The Museum Computer Network software is being adapted for use on the computer of the University of Arkansas.

2.3.27

Vance (1970) examines the GRIPHOS macro language for general humanities research in the light of existing museum records and the information needs of museums and their scholarly public. He describes the PL/I programs for the IBM 360/30 machine. The main difficulties are found to be the absence of a proofreading facility prior to input, the absence of lower case characters on the IBM 029 card punch, and the lack of core storage for large programs. Typical costs of computing

are reviewed. An improvement is the use of IBM 2741 terminal, which has full upper and lower case keyboard. The data is initially assembled on disc, then read to magnetic tape when sufficient has accumulated. Costs are found to be no greater than 90c/record, excluding bibliographies. Vance remarks that it remains to be seen whether the various museum projects will proceed independently or under the auspices of the Museum Computer Network.

2.3.28

Walker (1970, 2 references) gives a description of the prehistoric carvings and full data for the information retrieval exercise described by Wilcock (1970b). The coding procedure will be given in paragraphs 2.6.2 and 2.6.6. Distribution maps are by-products of the information retrieval runs and several of these, produced by programs developed for this thesis, are given in Walker's references.

2.3.29

Wilcock (1970a) describes a plain-language keyword-control information retrieval system, with logical tests of any complexity (using ALGOL 60 facilities). Distribution maps are drawn as a by-product. The program is fully described in paragraph 2.6.6.

2.3.30

Reynolds describes CODIL, a COntext-Dependent Information Language in four papers (1970, 1971a, 1971b, 1971c). The limitations of conventional information retrieval techniques are found to be a consequence of the "stored-program computer". The proposed system has variable-length data items of the form NAME = SMITH, YEAR = 1945 and DISEASE NOT MEASLES i.e. each has a set name, an operator and a value. A statement is a list of items separated by commas and terminated by a full stop. The level of an item is a number representing the distance of the item from the beginning of the list. Successive

statements may be combined by eliminating common leading items from one of them. This system allows rapid reading, writing and processing with no look-ahead and the minimum of back-tracking. During retrieval the CODIL system has a *facts statement* which is compared with all other statements (called the *criteria file*). If all facts in the facts statement are *true* in the criteria file statement, then any remaining items in the criteria file statement are transferred to the facts statement. A pilot program has been written for the ICL System 4-50 machine with CODIL pages of 1K bytes. The decision-making routine is less than 1K bytes in length. Processing efficiency is good, keeping the line printer operating at about 80% of its designed maximum speed. Disc accesses are the only other peripheral operations, and CPU utilisation is about 10%. The search is carried out at about 3,000 lines per second. Logical operators available are =, NOT, >, >=, < and <=. ANY and NONE are used to denote the complete set and the empty set. OR ABSENT is used to handle situations where the information is undefined. A data name by itself is used as a label. PRINT and TERMINATE RUN are examples of action statements. OR is implicit in a sequence of items at the same level, AND in connections between different levels. The ELSE connective is also allowed but must be explicitly stated. The use of this language effectively changes the "stored program computer" (in which information retrieval systems are either simple, requiring distortion of data, e.g. coding, abbreviation; or complex, expensive, difficult to alter and awkward in use) into the "information language computer" which is flexible and without restrictions. In the third 1971 reference an example of the use of CODIL in handling cave records is given.

2.3.31

Chenhall (1971a) reports the setting up of a committee by the

Archaeological Data Bank Conference, held 3-4 May 1971 at the University of Arkansas, to assemble the details of a minimal information system for archaeology, and to establish compatible data banks in various locations. The conference established, however, that compatibility was not absolutely essential between data banks provided there are no logical inconsistencies in the recorded data. The limitations are no longer in the hardware and software; the main problem now seems to be to decide what is to go into the system. The composition of the committee is:

Gundlach (Germany)

Lityak-King (Mexico)

Lyons (National Park Service, U.S.A.)

Reeves (Canada)

Voss (National Museum, Copenhagen)

Wagner (Waterloo Lutheran University, U.S.A.)

under the chairmanship of Chenhall. British workers (Lewis, Williams, Cutbill, Clark, Laflin, Wilcock) have been associated with the work of IRGMA, whose format has been accepted by UNESCO-ICOM. It needs only the acceptance of the IRGMA format by the above Archaeological Data Bank Conference committee for the possibility of truly international communication in archaeological records to become a reality.

2.3.32

Chenhall (1971b) continues to advance his ideal of a world-wide computer-oriented data bank for archaeologists. He states that computer hardware and software have only recently become adequate, organisational facilities become available, and archaeological theory become sufficiently defined for such a project to be feasible. The use of the remote terminal is described, and three operational projects:

Smithsonian, Museum Computer Network and GIPSY (University of Oklahoma, mainly work on N.American Indian material). The Smithsonian system is highly structured, and probably not flexible enough for archaeological material, but it has the useful feature of a *Global Reference Code*,

based on latitude and longitude. Areas may be specified and then tagged by a mnemonic. The Museum Computer Network and GIPSY systems are both unstructured, allowing the user to produce a hierarchy at the time of retrieval. Chenhall is of the opinion that standardisation is necessary before progress can be made. The Arkansas Archaeological Survey, which uses the Museum Computer Network Series of programs, is described. A magnetic tape Selectric typewriter is used for input. This allows records to be easily checked before transmission, easily amended and easily transmitted in magnetic tape form. Insertions can also be made easily. A description of the standard input form and codes are given, and the need for thesauri to avoid synonyms emphasized. A plea is made for standardisation before large bodies of material are converted for computer input.

Schneider (1971) disagrees with Chenhall, expressing the opinion that it may not be necessary or even desirable to implement a standard system for all data banks for museum and archaeological purposes. This assertion largely arises from the differing requirements of various users.

Vance (1971) makes several points in answer to Schneider. He does not suggest that one package will serve every need, but transcription of data is expensive, and so it is desirable that compatibility be achieved as soon as possible. Individual museums will each be responsible for the type of information collected and the manner in which it is recorded, but groups are urged to collect and record data uniformly. The Museum Computer Network system is mostly used by art museums, but it has been used for bibliographic records, archaeological records and excavation site records. It has several hundred data categories, but only about 20 are used for any specific application.

2.3.33

Vance and Heller (1971) give a description of the file system of the Museum Computer Network developed between 1967 and 1970 at New York University and continued at the State University of New York, Stony Brook. Lengths of files, records, segments and fields are variable and unlimited. The sequence of files, of records within a file, and of segments within a record are all random. Access is gained by multiple inverted indexes. Different types of file are available for object descriptions, biography, publications, events and places. The items on each file consist of a unique identifier and any number of indexed descriptions, with associated free text. An internal thesaurus is necessary to control synonyms. Each annotation within an item carries a type number from a standard list; this technique has been used in most of the information retrieval investigations carried out for this thesis, the most developed example being the Roman inscription scheme (see paragraphs 2.6.2 and 2.6.6). A *global reference code* may be used to specify any area in the world. There is a possibility that the Museum Computer Network scheme will be taken up by the Bodleian Library, Oxford.

2.3.34

Gaines describes the use of a remote terminal at the Navajo Indian Reservation in two 1971 references. The terminal was connected by telephone to a Honeywell 225 computer at Arizona State University, some 300 miles distant, and used to record and analyse data gathered by Field School students. Experiments were carried out on the following topics:

- (a) provision of on-line computer facilities at a remote site;
- (b) provision of the results of the excavation on a daily basis, with immediate analysis;

- (c) use of non-computer personnel to prepare input data and isolate errors;
- (d) design of a special language to enable field workers to use sophisticated computer techniques without the aid of computer personnel;
- (e) test and evaluate the programs used and developed under these circumstances;
- (f) demonstrate that the use of on-line computer facilities in a remote location really does aid the archaeologist in decision making, interpretation and verification.

During the Field School software was developed for format checking (with error reports), for file-building routines, and for search and analysis, with statistical checks and logical information retrieval. All these facilities are available in the suite of terminal programs developed for Doncaster Museum and the Department of the Environment (see paragraphs 2.6.12 - 2.6.17). A remote terminal was also operated at the University of Keele 1971 Summer School in Archaeology (lecturers Celoria and Wilcock), connected to the North Staffordshire Polytechnic ICL System 4-50. None of the students were found capable of writing their own archaeological programs, but several were able to use packages already developed with worthwhile results. The exercise is to be repeated in 1972.

2.3.35

Hebditch (1972) describes a terminal language for the administration of a data base. The syntax is based on the following commands:

find print alter new move why end

The reply to find x is the block header of x, to confirm the identification of the item. The reply to print y is the appropriate name of the sub-item and the value of the sub-item. The reply to alter y is

the name and the old value, plus a request to type the new value. For new x, the record number x is checked to be vacant, then a request is made for the new item to be typed in, sub-item by sub-item, with sub-item names. The command MOVE concerns the transfer of items from one group of records to another. The command end causes logoff. The command why is a request for the clarification of error messages. Several of these features are similar to those offered by the terminal suite of information retrieval programs developed for Doncaster Museum (see paragraphs 2.6.12 - 2.6.17). Equivalent facilities to the find/print, alter, new and end commands above are available.

2.3.36

Chouraqi (1972) reports FORTRAN IV software for the IBM 360 machine for bibliographic search applied to Roman archaeology.

2.3.37

Miss S.Laflin (Mrs P.Barker) of the Computer Centre, University of Birmingham, has developed KDF9 software for the archaeological gazetteer of Shropshire. "*Generic types*" are used to describe artefacts, e.g. amber, bone, industrial site, etc. each with several possible sub-types. Bibliographic references are also entered. A code-sheet is necessary for the interpretation of a record. The input is by punched cards, in fixed format. This system has little flexibility, and the stylised coding is a severe disadvantage, although steps have now been taken to amplify the codes before printout.

2.4 Roman Inscriptions

2.4.1

The Roman invasion of Britain and the occupation which lasted four centuries has left a wealth of inscriptions which serve to document the everyday lives and deaths of soldiers and ordinary people. If this body of information could be made the subject of modern information retrieval methods, new facts concerning the movements of troops and individuals, the history of forts and civil settlements and changes in epigraphic style might be discovered. The object of this study has been to show that such a scheme is possible. However, copyright law has dictated that the magnetic tape version of the prime reference work *The Roman Inscriptions of Britain* (Collingwood and Wright, 1965) be regarded as a "single copy for research purposes only". The direct transcript, with the exception of the Ribchester inscriptions RIB 583-599, 2268 and 2269 (for which special permission has been obtained from the publishers and author), cannot be published, although any new interpretations brought to light may be published. A full description of the classification scheme which has been evolved is given in paragraphs 2.6.2 and 2.6.6. There are four main types of inscriptions to be classified: religious dedications, tombstones, honorific inscriptions and building inscriptions.

2.4.2

The first comprehensive publication of the Roman inscriptions found in the British Isles was published in 1873 by Hübner as volume vii of the *Corpus Inscriptionum Latinarum* (CIL). It contains 1200 inscriptions on stone, about 100 on metal, 90 on pottery and many potters' stamps. Hübner makes some glaring errors in geography, placing the Mendips in Derbyshire and Denbighshire in Scotland, but despite these mistakes Hübner's work is a remarkable feat for its

time. Hübner (1877, 2 references; 1881) continued to publish additions to CIL in *Ephemeris Epigraphica* (EE), but the work never received a combined index to enable scholars to trace inscriptions easily.

2.4.3

The fourth and fifth additions to CIL were continued by Haverfield (1892,1913) in *Ephemeris Epigraphica* until the termination of that periodical.

2.4.4

Another reference work, edited by Dessau (1892-1916) is *Inscriptiones Latinae Selectae* (ILS) which appears in five parts.

2.4.5

After the termination of *Ephemeris Epigraphica* Haverfield (1913,1914) published continuations of the additions to CIL and EE in British Academy supplemental papers, the prototypes of the summaries of inscriptions from Roman Britain edited by Collingwood between 1921 and 1938 in the *Journal of Roman Studies*. Haverfield had hoped that G.L.Cheesman, author of *The auxilia of the Roman imperial army* would undertake a fresh publication of the Roman inscriptions of Britain, but unfortunately he was killed in action in 1915. Haverfield then made plans to launch a scheme for the editing and publication of the new work, to be styled *The Roman Inscriptions of Britain* (RIB). Haverfield decided that the stones and their texts, whether clear or indistinct, were to be published as line drawings, and he chose Collingwood to perform this task. Haverfield died in 1919, bequeathing his books to the Ashmolean Museum and making the University of Oxford his residuary legatee. The administrators of his bequest made the editing and publication of the future work a primary claim on their funds. Collingwood continued the work until

1938, when a breakdown in health, brought on by the excessive work which he was undertaking, forced him to take on additional help. His choice for Junior Editor was R.P.Wright.

2.4.6

Wright (1939-1970) has continued the work of Collingwood in the publication of annual summaries of new inscriptions in the *Journal of Roman Studies* (JRS). He has also brought to completion the publication of Vol.I of RIB (1965). This volume contains 2400 inscriptions, mostly on stone, although a few are on metal. Some of the inscriptions are milestones, Roman inscriptions imported from other countries which have been mistaken for Romano-British stones, genuine inscriptions of other periods which have been mistaken as Roman, forgeries, and marked stones. For this publication deadlines were set at 31 December 1954 for inscriptions on stone and at 31 December 1956 for other items, all items discovered after these dates being excluded. The work has concordance tables giving the correspondence between reference numbers in CIL, EE, ILS, Roman Britain in 1913, 1914 (Haverfield), JRS and RIB. There is also an index of place names, but no other cross-indexes. The ordering of the inscriptions commences in London (LONDINIUM) and ends with Roman Scotland but apart from a general south to north direction there seems to be no logical order. The lack of general cross-indexes is a severe disadvantage, and this is a problem which the computer could easily solve. Indexes are in the course of preparation by hand and Wright in personal correspondence states that these are about half completed (1972), but the suggestion that the computer could be used to assist has not met with any sympathetic response. Duncan Cameron (National Director of the Canadian Conference of the Arts, Coordinator of the ICOM sub-committee on Public and Modern Art) in *(UNESCO) Museum* 23(1), 1970/71 suggests

that the main reason for mistrust of computers by academics is that the analysis is taken out of their hands. The removal of boring and wasteful tasks by use of the machine also unfortunately leads to a loss of personal control of the project. The analysis by computer, reducing a lifetime's work to a few hours or days also leads to a loss of personal prestige, or so it is thought by some academics; they therefore stick to their old methods, even though this may well deprive their own generation of the fruits of their labours, and they may even die with the task incomplete (who then is to know what their conclusions were to be?). Perhaps the case of the RIB index is just such a situation as Cameron describes.

2.4.7

Leo Rivet (1965), compiler of the Ordnance Survey maps of Southern Britain in the Iron Age and Roman Britain has produced notes on Roman epigraphy for use in the Department of Classics, University of Keele. The main types of inscriptions are religious dedications, honorific inscriptions, building inscriptions, milestones, legal enactments and decrees, tombstones, military diplomas and miscellaneous (ingots of metal, seals, potters' stamps, votive plaques, oculists' stamps and graffiti). A list of the 35 voting tribes is also given, with their common abbreviations or inscriptions.

2.4.8

Jory (Jory & Moore, 1966; Jory, 1968) reports MACRO 6 (DEC) software for the PDP-6 machine equipped with magnetic tapes and remote teletype, for the purpose of producing an index to CIL vi, which covers about 40,000 inscriptions originating in Rome. The data is punched on paper tape, transferred to magnetic tape then edited and sorted by computer. There are problems in producing machine-readable form, particularly for Greek letters, cognomina and Roman numerals

(which must be differentiated from letters). Algorithms are being developed for the recognition of features in the text. The goal is the establishment of a word index and analysis of dating criteria with a view to further investigation and historical research.

2.4.9

The conference on Roman Epigraphy, 1967, made suggestions for the coding of inscriptions. It was proposed that the coding for each inscription should be fitted onto a single 80-column card. The suggested coding scheme goes into great detail on some points, e.g. occupation, relationship, while other important classifications, e.g. map reference, are omitted entirely. This scheme suffers from the disadvantages of all numerical coding schemes in that the codes must be learnt or looked up, and the system is not attractive to archaeologists and scholars lacking computer training. The coding scheme proposed in this study (see paragraphs 2.6.2 and 2.6.6) allows plain-language alphanumeric descriptions to be input, and as many cards as necessary may be used for each inscription.

2.4.10

Thompson (1968) describes the use of the *pelta*, or light shield as a motif at the extreme left-hand and right-hand edges of inscriptions. The terminals of these peltae are often carved with bird or griffin heads. Examples are given from RIB.

2.4.11

Oehler (1970/71) discusses the electronic documentation of a collection of Roman sculpture photographs, comprising 45,000 photographs covering 28,000 individual Roman sculptures. The computer classification is carried out by Gundlach (1970/71) under 104 headings, each named by a keyword. Examples of the classification headings are museum, place of discovery, size, parts of the sculpture preserved,

sex of the subject, clothing, ornaments, gods, insignia, dating, workshop, and publications. Inscriptions are also described under such headings as principal name of the subject, location of the inscription on the statue, name of the artist, other names, and alterations to the text. The treatment is not as detailed as that proposed in this study. The name of the Oehler and Gundlach project is *Monumenta Artis Romanae*.

2.4.12

Stefan (1971) describes the application of mathematical methods to epigraphy, investigating methods of measuring the degree of resemblance between different character styles of the same script. In the body of data analysed, Greek inscriptions of the Roman period from Histria in Dobruja, each character style was used only for a limited period. A few dozen inscriptions dating from the 3rd century B.C. to 2nd century A.D. were analysed to give incidence matrices based on presence/absence of each variable rather than percentage occurrence. The method of Hole and Shaw using a *discordance measure* was used to seriate the matrices. Further sets of data were a set of Greek inscriptions of the Classic, Hellenistic and Roman periods; and a set of Greek inscriptions from the Late Hellenistic and Roman periods. The variables used are the different forms of the full characters, each taken as an entity. No attempt is made to classify the characters in terms of parts of characters or numerical representation of curves.

2.4.13

Chouraqui (1972) reports FORTRAN software for the UNIVAC 1106 machine to process CIL viii. The aim is to automatically update and edit the data, produce indexes and tables and elaborate a documentation system. CIL viii contains 800 inscriptions concerning veterans

in Roman Africa. Codes are being derived for the description of texts (reading, interpretation), peculiarities of the language (morphology, linguistic particulars), context (present location, material, size, ornamentation, conservation) and semantics. This project has similar aims, but is wider in scope than the CIL vi project of Jory and Moore (see paragraph 2.4.8). It is indeed a pity that the attempt to mount a similar project for CIL vii (Roman inscriptions of Britain) has not been supported by the publishers and surviving author of the British reference work RIB.

2.5 *Ribchester Museum Catalogue*

The Roman fort of BREMETENNACVM, guarding the crossing of the River Ribble for the main west coast Roman road, has been of antiquarian and archaeological interest for two centuries. Its main claim to fame is the "Ribchester helmet" now in the British Museum. The site is covered by the Parish Church of Ribchester, the Ribchester Museum, and adjoining houses. The south-eastern corner of the fort has been eroded away by the Ribble, and the only visible remains are the foundations of granaries behind the present museum. The museum itself was founded by Miss Margaret Greenall soon after the turn of the century. It has never had a catalogue. The contents of the museum are to be filed on the computer, sorted and printed. The line printer output is to be reproduced by offset litho and sold to the public. There will be sections, when complete, for inscriptions, coins, pottery, glass, metal, miscellaneous objects, models and a bibliography. At present only the inscriptions and bibliographic sections are complete, the remainder awaiting identification by various authorities, or routine archaeological analysis and labelling. Once created, the file will be updated when necessary, and a new edition of the catalogue can be run off at any time. The inscriptions comprise RIB 583-599, 2268 and 2269 (for which special permission to publish has been obtained from Oxford University Press and R.P. Wright), and the bibliography which has been compiled is given at the end of the chapter.

2.6 *Information Retrieval Programs developed*

2.6.1

ARCJW005 Pottery volumes (ALGOL 60)

This program is designed to calculate the volume of a wheel-made pot, the inner profile of which is recorded on a d-Mac pencil follower. The volume of a pot is a routine statistic which is necessary for an information retrieval and classification scheme for pottery. The program is fully described in the Pottery Classification chapter, paragraph 4.4.1.

2.6.2

ARCJW006 Data Transcriber with Magnetic Tape and Line Printer Output (ALGOL 60)

This program has been used for all information retrieval studies in this thesis. Its purpose is to transcribe records from cards and paper tape to magnetic tape. There is also a listing on the line printer. The program first determines whether the run is to be in the graphics mode or not. The designated magnetic tape is then prepared. A series of cards comprising a record is read. The end of the record is indicated by a change in the record number (punched in columns 73-80), and the record is then written to magnetic tape. A card counting feature allows the operation to reload the card reader during the reading of a large deck of cards, a RELOAD message being given every 500 cards. When the run is in the graphics mode a d-Mac paper tape is also read, the coordinates rotated, smoothed and scaled to a standard height with adjustment of scale factor, and the scale factor recorded on magnetic tape followed by the processed coordinates. There is a check that the identifiers of the card and paper tape records tally. When all card and paper tape records have been read, the magnetic tape is rewound, a count of

the number of records is output, then the magnetic tape is re-read from the beginning. In the graphics mode any figures present are plotted on the digital incremental plotter and labelled with the item number and the scale. Card images are printed on the line printer. The program used procedures DECINT, INCARD, OUTCDI, INSBL, OUTSBL and INCDI, which are described fully in Appendix A. Typical uses of the program will now be presented with their coding schemes.

2.6.2.1

Roman inscriptions

This application allows plain-language entries in alphanumeric, each record using as many cards as desired. All cards carry the record number, in this case the RIB reference number of the inscription, in columns 77-80, right justified. Each card carries a type designator in columns 2 and 3, with column 1 reserved for a repetition factor. The type designator indicates what sort of information is to be found on the remainder of the card. The repetition factor is left blank for the first or only card of any type. Subsequent cards of the same type are punched with 1,2,3....8 in column 1, allowing up to 9 cards for each type of information.

The layout of the cards is as follows:

<i>Type Desig- nator</i>	<i>Contents</i>	<i>Columns of field</i>	<i>Comments</i>
01	Roman name of site	11-40	
	Modern name of site	41-70	
02	Modern county	11-40	
	Roman grouping (e.g.HADRIAN'S WALL)	41-70	
03	NGR of site	11-20	e.g.ST 1234
	Width of stone	21-30	
	Height of stone	31-40	
	Depth of stone	41-50	e.g.JDW YYMDD e.g.671205 for 5 December 1967
	Initials of classifier	51-60	
	Date of classification	61-70	
04	General description of stone	11-40	
	Material	41-70	
05	Present location	11-40	

06	Detailed location or site	11-40
	" " " "	41-70
07	Associated finds	11-40
	" "	41-70

nlx, n2x refer to the object of dedication, e.g. God, buried person, etc.

10	Praenomen	11-30	
	Nomen	31-50	
	Cognomen	51-70	
13	Filiation	11-30	
	Tribe	31-50	
	Origo	51-70	
15	General Origo	11-30	e.g.Jewish
17	Status	11-40	e.g.Libertus, Augustus
	Occupation	41-70	e.g.Miles, Emperor
19	Sex	11	
	Social relationship to dedicator	21-40	
21	"Virtues"	11-30	e.g.Pientissima
	"	31-50	
	"	51-70	
22,23	Career details	11-40	e.g.Pontifex
		41-70	Maximus, Cos, Trib. Pot., Caesar, Pater Patriae, Legatus
24	Age	11-16	YYMDD
	General Age Term	21-30	e.g.Child
	Military service (years)	31,32	
	COS number	41,42	
	TRIB POT number	51,52	
27	Name of deity, or D M etc (not abbreviated)	11-40	
		41-70	
28	Cases following D M	11-20	
		21-30	
		31-40	
		41-50	
		51-60	
		61-70	

n3x, n4x serve the same purpose for the dedicator as nlx, n2x do for the object of the dedication

30	Praenomen	11-30
	Nomen	31-50
	Cognomen	51-70
33	Filiation	11-30
	Tribe	31-50
	Origo	51-70
35	General Origo	11-30
37	Status	11-40
	Occupation	41-70
39	Sex	11
	Social relationship to object of dedication	21-40
41	"Virtues"	11-30
		31-50
		51-70

42,43	Career details	11-40 41-70	
44	Age	11-16	
	General Age Term	21-30	
	Military service (years)	31,32	
47	Motive for dedication	11-40 41-70	e.g. V S L M
50	Date of inscription or generalised date	11-40	e.g. 112 < HADRIAN
	Military units	41-70	e.g. LEG XX VV
51	Motifs	11-40 41-70	
52	Curses	11-40 41-70	
53	Unusual spelling	11-40	actual (usual)
	Unusual punctuation	41-70	e.g.leaf
54	Unusual features	11-40 41-70	
55	Number of persons object	11-40	e.g.2, or FILII
	Number of persons dedicating	41-70	
57	= 1 if fragmental	11	
	= 1 if the Emperor is the object of the dedication	21	
	= 1 if reused	31	
	= 1 if evidence of erasure	41	
	= 1 if authentication doubtful	51	
59	Formulae	11-30 31-50 51-70	e.g. H S E D D
63	Grandson of	11-40	
	Place for freedman reserved (formula)	41-70	

n8x, in order 80-89, 180-189 etc.

Actual inscription on stone as in RIB
including omissions and reconstructions.

A typical printout of a Roman inscription record is given in Figure 2.1

2.6.2.2

Petroglyphs

This application differs from the Roman inscriptions scheme in that a limited amount (one card per record) of alphanumeric coding is used. Subsequent cards of the record are in plain language. Each record may have as many cards as desired. All cards carry the record number, in this case Dr M.J.Walker's reference number (see Walker, 1970, An analysis of British petroglyphs; Wilcock, 1970, Petroglyphs by computer) in columns 71-80, right justified. Each card carries a type designator in columns 2 and 3, with column 1 reserved for a repetition factor. The

01	BREMENNACVM	RIBCHESTER	RIB	583
02	LANCASHIRE	REGIO BREMENNACENSIS	RIB	583
03	SD6535 24 51 20 JDW 691109		RIB	583
04	SHAFT OF RECTANGULAR PEDESTAL		RIB	583
05	RIBCHESTER MUSEUM		RIB	583
51	RELIEF OF APOLLO	CLOAK	RIB	583
151	PHRYGIAN CAP	QUIVER	RIB	583
251	LYRE	FEMALE FIGURES	RIB	583
351	TURRETED CROWN	VEIL	RIB	583
30	AELIUS ANTONINUS		RIB	583
33		MELITENE	RIB	583
37	PRAEPOSITUS NUMERI ET REGIONIS	LEG VI VICTRICIS	RIB	583
39	M		RIB	583
42	PRAEPOSITUS NUMERI ET REGIONIS		RIB	583
47	WELFARE OF EMPEROR	WELFARE OF UNIT	RIB	583
17	DOMINUS NOSTER		RIB	583
19	M		RIB	583
27	APOLLO	MAPONUS	RIB	583
50	>GORDIAN	NUM EQ SARMATARUM BREM GORD	RIB	583
150		LEG VI VICTRIX	RIB	583
127	REGIO BREMENNACENSIS	BRITANNIA INFERIOR	RIB	583
55	1	1	RIB	583
80	DEO SAN(CTO)/[A]POLLINI MAPONO/[PR]IO SALUTE D(OMINI) N(OSTRI		RIB	583
81)/[ET] N(UMERI) EQ(UITUM) SAR/[M(ATARUM)] BREMENN(ACENSIVM		RIB	583
82)/[G]JORDIANI/[A]EL(IUS) ANTONI/NUS C(ENTURIO) LEG(IONIS) VI/		RIB	583
83	VIC(TRICIS) DOMO/MELITENIS/PRAEP(OSITUS) N(UMERI) ET R(EGION		RIB	583
84	IS)/.../.../.../.../...		RIB	583

Figure 2.1

Typical Roman inscription printout

type designator indicates what sort of information is to be found on the remainder of the card. The repetition factor is left blank for the first or only card of any type. The layout of the petroglyph cards is as follows:

<i>Type Designator</i>	<i>Contents</i>	<i>Columns of field</i>	<i>Comments</i>
10	R (for rock) plus one of the following: R (natural rock surface) S (slab) I (integrated rock) M (menhir) P (portable stone) Blank (uncertain)	11,12	Coded card
	S (for site) plus one of the following: M (megalithic monument) S (stone circle) C (cist) B (barrow) O (occupation or industrial site) Blank (uncertain)	14,15	
	A (for age) plus one of the following: N (Neolithic) B (Bronze Age) I (Iron Age) Blank (uncertain)	17,18	
	H (for height) plus a single digit representing the height above sea level as follows: 1 (0-250 ft) 2 (250-750 ft) 3 (above 750 ft)	21,22	
	D (for distance) plus a single digit representing the distance from the nearest navigable water as follows: 1 (under $\frac{1}{2}$ mile) 2 ($\frac{1}{2}$ - 5 miles) 3 (5 - 10 miles) 4 (10 - 30 miles) 5 (over 30 miles)	26,27	
	C (for class of carving) plus one or more digits, each specifying a class of carving, in the range 1-8 as follows:	31-39	e.g.C1258

- 1 (cup marks)
- 2 (cups and rings)
- 3 (spirals and labyrinths)
- 4 (crosses, swastikas, and segmented circles)
- 5 (rings, ellipses, hooks and loops)
- 6 (linear patterns)
- 7 (carreg saethau)
- 8 (weapons, feet, animals and people)
- Q (questionable)

M (for magnitude) plus one of the following: 41,42

- 1 (single stone)
- 2 (multiple site)

National Grid Reference 51-58 e.g. SJ123156
 consisting of two letters and six figures, or Irish grid reference consisting of one letter and six figures (in the latter case column 51 is left blank)

Initials of classifier 61-63 e.g. JDW

Date of classification in 65-70 e.g. 680523
 format YYMMDD, allowing dates for 23 May 1968
 to be treated as numbers for ease of comparison.
 This enables records made prior to a certain date to be modified in the light of new information.

Item classification number 71-80

20 Name of site 11-40
 (site names)

Name of parish 41-70

Item classification number 71-80

30 Author and date 11-30
 (Bibliography)

Journal (abbreviation) 31-40 e.g. ANT J
 (Antiquaries Journal)

Volume number 41-50

Pagination 51-60

OR

Book 31-60

Pagination 61-70

OR, for Royal Commission on Ancient
and Historical Monuments literature

RCAM, RCAMS etc. 31-40

County 41-55

Number 56-60

Pagination 61-70

Item classification number 71-80

40 Detailed location on site 11-70
(Detailed site
information) Item classification number 71-80

50 Associated find 11-70
(Associated
finds) Item classification number 71-80

Part of a typical petroglyph listing is given in Figure 2.2

2.6.2.3

Clay pipes

This application is principally concerned with the GRAPHICS mode of operation of the program. Profiles of clay pipes are normalised and stored on magnetic tape. Typical features recorded in card image form might be reference number, date range, and distinctive stylistic points (e.g. heart-shaped base, milling, degenerate spur) for the bowl of the pipe. It may also be an advantage to record length and diameter of stem where known, and bore diameter (although this last feature has been shown to be very variable by Dr F.Celoria and therefore probably of little diagnostic use).

A typical graphical output for a clay pipes run is given in Figure 2.3. Clay pipe data has been taken from Atkinson,D. and Oswald, A., London Clay Tobacco Pipes.

BRONZE AGE BARROW SITES
WITH CUPS OR CUPS AND RINGS

The Result of a Computer Information Retrieval Exercise.
This is the output produced on the line printer.

10	RS SB AB H2 D4 C2	M1	N0484529 JDW 691230	073
20	BLACKGATE	RESCOBIE		073
30	SIMPSON (1867) ASCC	68		073
10	RS SB AB H2 D4 C1	M1	NZ006853 JDW 700101	205
20	FAWNS	KIRK WHELPINGTON		205
30	GREENWELL (1877) BB	342FF.,433		205
10	RS SB AB H3 D4 C1	M2	NY895793 JDW 700101	206
20	PITLAND HILLS	CHAPELRY OF BIRTLEY		206
30	HALL PSAN	12 270		206
130	HODGSON (1897) HIST. OF NORTHUMBERLAND	VOL.4,353		206
10	RS SB AB H2 D2 C126	M2	NZ695189 JDW 700103	231
20	HOWE HILL	LOFTUS		231
30	HORNSBY & STANTON YAJ	24 267		231
10	RS SB AB H2 D1 C146	M2	NZ804170 JDW 700103	233
20	HINDERWELL BEACON	HINDERWELL		233
30	HORNSBY & LAVERICK YAJ	25 445-447		233
10	RS SB AB H2 D2 C1	M1	NZ833147 JDW 700103	235
20	WHINNY HOWE, GOLDSBOROUGH	LYTHE		235
30	GREENWELL ARCH	52 44-45		235
10	RS SB AB H2 D2 C2	M2	SE977998 JDW 700103	237
20	RUDDA HOWES	STAINTONDALE		237
30	TISSIMAN ARCH	36 446-449		237

Figure 2.2
Petroglyph
Listing

2.6.3

ARCJW007 Map plotter (for distribution maps) (ALGOL 60)

This is an experimental version of a program to plot a map outline and to add a series of symbols at specified map references to form a distribution map. The technique is of particular use in information retrieval exercises, where it is a by-product of the retrieval and listing of items satisfying given criteria, and is developed in programs ARCJW002 (paragraph 2.6.4), ARCJW017 (paragraph 2.6.7) and ARCJW011 (paragraph 2.6.6). The program is fully described in the Graphics chapter (paragraph 3.4.3).

2.6.4

ARCJW009 MAPCBI MAP of Great Britain and Ireland Test Program (ALGOL 60)

This is a skeleton program to test the procedure MAPCBI, a tailor-made routine which plots a map of Great Britain and Ireland complete with border, scale and grid north point. A distribution map may be plotted using specified grid references on the British National Grid and Irish Grid. The program is fully described in the Graphics chapter (paragraph 3.4.4).

2.6.5

ARCJW010 Amend or list magnetic tape files (ALGOL 60)

A flowchart for this program is given in Figure 2.4. The program commences by identifying the required old and new master tapes. An amendment card is read from the card reader, and a full record, consisting of several card images, from the old master. The card identifier (ci) is compared with the record identifier (ri), treating blank characters and zeros as equal. If $ci = ri$ then the record is copied unchanged to the new master tape; if $ci \neq ri$ the card is ignored; if $ci = ri$ the card code is checked against the codes of the card images

5

SCALE 0.9

4

SCALE 0.3

3

SCALE 0.3

2

SCALE 0.8

1

SCALE 0.3

Figure 2.3

Clay pipe diagrams

2.6.3

ARCJW007 Map plotter (for distribution maps) (ALGOL 60)

This is an experimental version of a program to plot a map outline and to add a series of symbols at specified map references to form a distribution map. The technique is of particular use in information retrieval exercises, where it is a by-product of the retrieval and listing of items satisfying given criteria, and is developed in programs ARCJW009 (paragraph 2.6.4), ARCJW017 (paragraph 2.6.7) and ARCJW011 (paragraph 2.6.6). The program is fully described in the Graphics chapter (paragraph 3.4.3).

2.6.4

ARCJW009 MAPGBI (MAP of Great Britain and Ireland) Test Program (ALGOL 60)

This is a skeleton program to test the procedure MAPGBI, a tailor-made routine which plots a map of Great Britain and Ireland complete with border, scale and grid north point. A distribution map may be plotted using specified grid references on the British National Grid and Irish Grid. The program is fully described in the Graphics chapter (paragraph 3.4.4).

2.6.5

ARCJW010 Amend and list magnetic tape file (ALGOL 60)

A flowchart for this program is given in Figure 2.4. The program commences by identifying the required old and new master tapes. An amendment card is read from the card reader, and a full record, consisting of several card images, from the old master. The card identifier (ci) is compared with the record identifier (ri), treating blank characters and zeros as equal. If $ci > ri$ then the record is copied unchanged to the new master tape; if $ci < ri$ the card is ignored; if $ci = ri$ the card code is checked against the codes of the card images

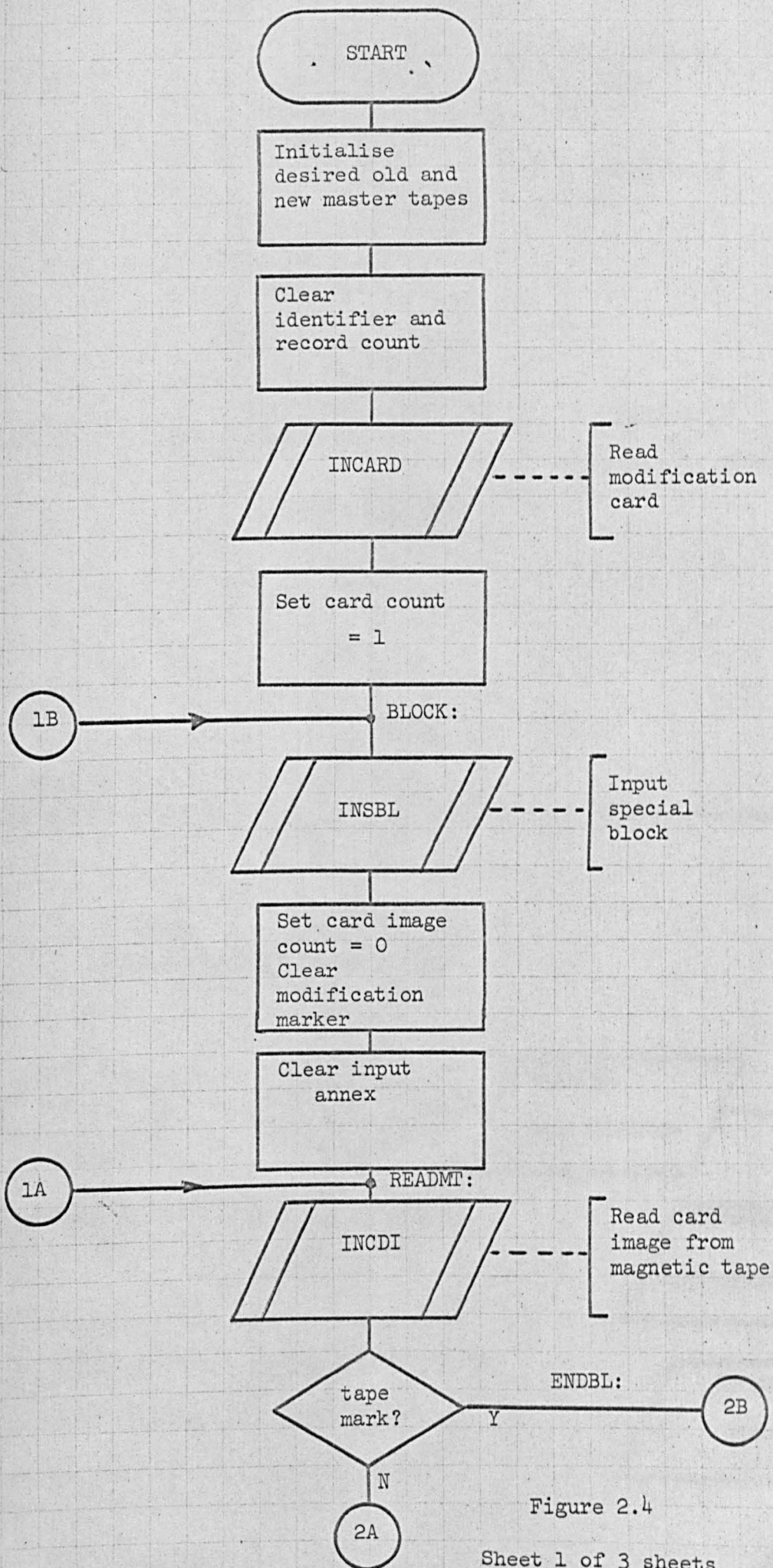


Figure 2.4

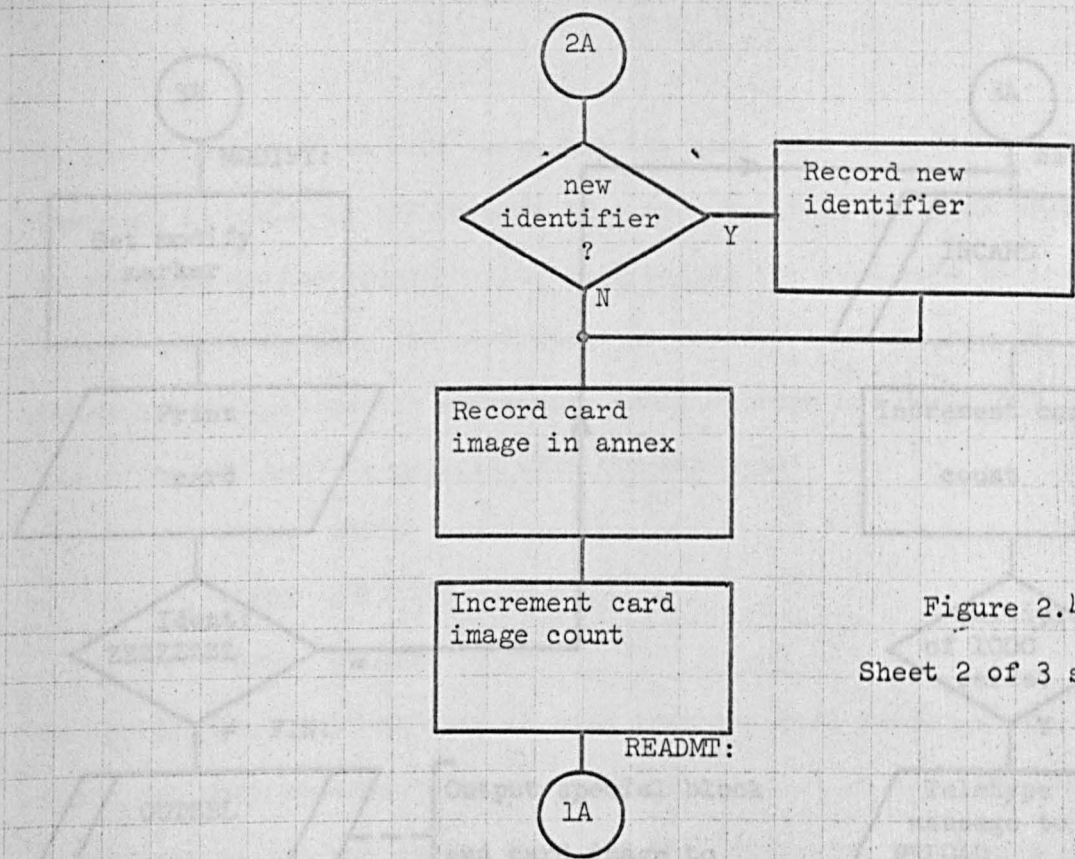
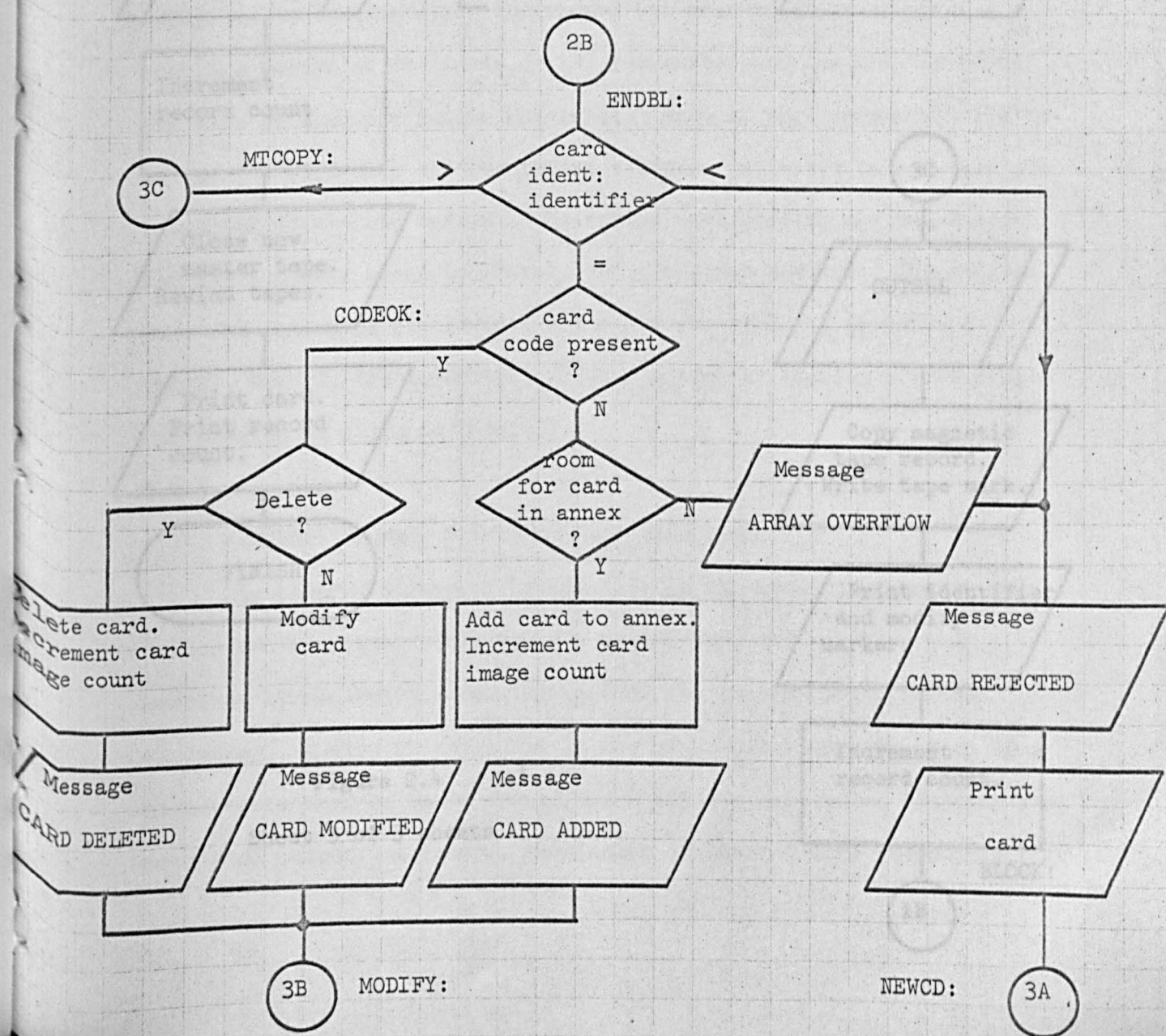


Figure 2.4

Sheet 2 of 3 sheets



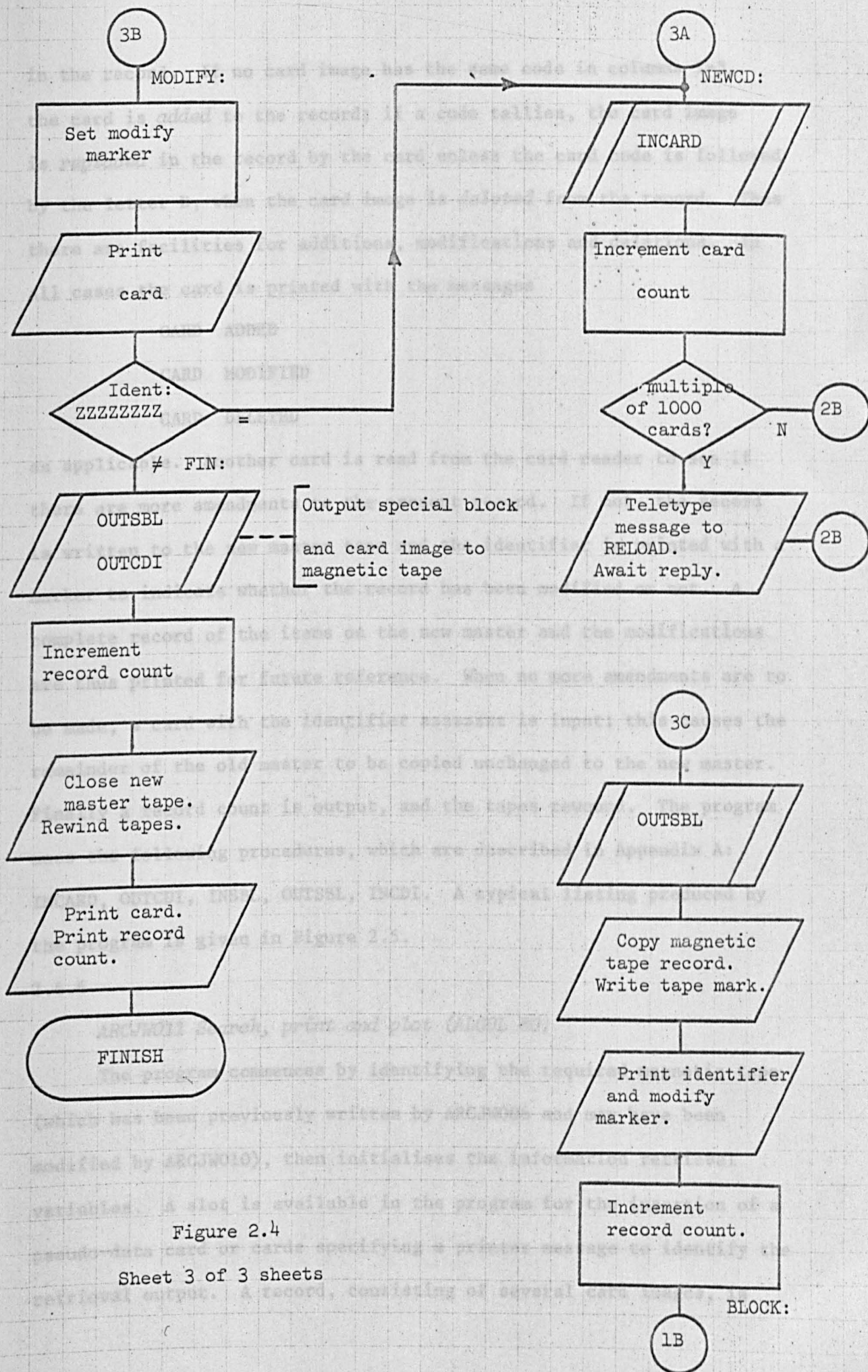


Figure 2.4

in the record. If no card image has the same code in columns 1-3, the card is *added* to the record; if a code tallies, the card image is *replaced* in the record by the card unless the card code is followed by the letter D, when the card image is *deleted* from the record. Thus there are facilities for additions, modifications and deletions. In all cases the card is printed with the messages

CARD ADDED

CARD MODIFIED

CARD DELETED

as applicable. Another card is read from the card reader to see if there are more amendments to the present record. If not, the record is written to the new master tape and the identifier is printed with a marker to indicate whether the record has been modified or not. A complete record of the items on the new master and the modifications are thus printed for future reference. When no more amendments are to be made, a card with the identifier zzzzzzzz is input: this causes the remainder of the old master to be copied unchanged to the new master. Finally a record count is output, and the tapes rewound. The program uses the following procedures, which are described in Appendix A: INCARD, OUTCDI, INSBL, OUTSBL, INCDI. A typical listing produced by the program is given in Figure 2.5.

2.6.6

ARCJW011 Search, print and plot (ALGOL 60)

The program commences by identifying the required magnetic tape (which has been previously written by ARCJW006 and may have been modified by ARCJW010), then initialises the information retrieval variables. A slot is available in the program for the insertion of a pseudo-data card or cards specifying a printer message to identify the retrieval output. A record, consisting of several card images, is

000			
001			
CARD MODIFIED			Modification
30		0002	
CARD MODIFIED			Modification
130		002	
CARD DELETED			Deletion
230D		002	
CARD ADDED			Addition
430		002	
002	M		
CARD MODIFIED	(Modification marker)		Modification
30		003	
CARD REJECTED			Rejection
130		002	
003	M		
004	(Modification marker)		
005			
006			
007			
008			
009			
010			
011			
012			
013			
014			
015			
016			
017			
018			
019			
020			

Figure 2.5

read from tape. If the run is complete (this is indicated by record identifier zzzzzzzz) the number of records which satisfied the information retrieval criteria is printed out, the tape is rewound, and if at least one satisfactory record has been retrieved the option is given for any map to be plotted (using DRAWMAP, see Appendix A) and for the retrieved map references to be plotted on it (using PLOTREFS, see Appendix A). Otherwise each record is checked against an optional boolean expression which is specified by pseudo-data cards consisting of ALGOL statements inserted in the slot provided. The logic checks may use ALGOL boolean operators *and*, *or* and *not*; relational operators *=*, *≠*, *<*, *≤*, *>*, *≥*; *switch*; the specially designed procedures SEARCH (checks for the presence of a specified alphanumeric string anywhere within a specified data area), DECINT (converts a string of numerical characters to an integer) and NOF (a boolean procedure which is assigned the value *true* if n of the m specified boolean variables are *true*); and any number of integer and boolean information retrieval variables. Thus a complex logical check is available by the simple artifice of using the ALGOL compiler, and this seems to be a more sensible design than the construction of a complex software routine for the interpretation of logic expressions. A typical logic check is given in Figure 2.6. If the check on any record succeeds, the record count is incremented, the grid reference is recorded in core ready for the plotting of the distribution map, and the record is printed in full. If the check fails, none of the above actions occur. The program uses procedures GETSTRING and SCRIPT (specify and produce packed strings including lower case alphabetic characters for the legends on distribution maps); SEARCH, DECINT and NOF (for use in the logic check); DRAWMAP and PLOTREFS (for the plotting of distribution maps); and INSBL and INCDI (for magnetic tape handling). Full descriptions of these procedures are

"COMMENT"MODIFIABLE BOOLEAN CHECK ON RECORD;

"FOR" I:=1"STEP"1"UNTIL"CDIMCT"DO"

"BEGIN"

"IF"DECINT(STORE,I*80-78,2)=10

"THEN""BEGIN""GOTO"CRIT1IRINT1;

CRIT1:"IF"STORE[I*80-65]=34"THEN""GOTO"OK"ELSE""GOTO"BLOCK;

CRIT2:"IF"STORE[I*80-65]=35"THEN""GOTO"OK"ELSE""GOTO"BLOCK;

CRIT3:"IF"STORE[I*80-65]=45"THEN""GOTO"OK"ELSE""GOTO"BLOCK;

CRIT4:IRINT2:=DECINT(STORE,I*80-3,4);"IF"SEARCH(STORE,I*80-48,8,B,1,1)

"AND"IRINT2"NE"75"AND"IRINT2"NE"182"AND"IRINT2"NE"220"AND"IRINT2"NE"238

"AND"IRINT2"NE"256"AND"IRINT2"NE"258"AND"IRINT2"NE"263"AND"IRINT2"NE"321

"AND"IRINT2"NE"143"AND"IRINT2"NE"154"THEN""GOTO"OK"ELSE""GOTO"BLOCK;

CRIT5:"IF"STORE[I*80-68]=48"THEN""GOTO"OK"ELSE""GOTO"BLOCK;

CRIT6:"IF"SEARCH(STORE,I*80-48,8,B,2,1)"THEN""GOTO"OK"ELSE""GOTO"BLOCK;

CRIT7:"FOR"J:=1"STEP"1"UNTIL"4"DO"BOOL[J]:=SEARCH(STORE,I*80-48,8,B,

2+J,1);

"IF"NOF(1,4,BOOL)"THEN""GOTO"OK"ELSE""GOTO"BLOCK;

CRIT8:"FOR"J:=1"STEP"1"UNTIL"4"DO"BOOL[J]:=SEARCH(STORE,I*80-48,8,B,

2+J,1);

"IF"NOF(2,4,BOOL)"THEN""GOTO"OK"ELSE""GOTO"BLOCK;

OK:SAT:="TRUE";"GOTO"SUCC;

"END";

"END";

"COMMENT"END OF MODIFIABLE BOOLEAN CHECK;

"GOTO"BLOCK;

SUCC:GRRECCT*3+2):=("IF"STORE[I*80-29]=0"THEN"3"ELSE"STORE[I*80-29]);

GRRECCT*3+3):=STORE[I*80-28];

GRRECCT*3+4):=DECINT(STORE,I*80-27,6);

RECCT:=RECCT+1;

LIST:"FOR" I:=1"STEP"1"UNTIL"CDIMCT"DO"

"BEGIN"

"FOR" J:=1"STEP"1"UNTIL"80"DO"

CHAROUT(4,STORE[I*80+J-80]);

"PRINT"PUNCH(4),'L';

"END";

"PRINT"PUNCH(4),'L2';

"GOTO"BLOCK;

Criterion 1

2

3

4

5

6

7

8

Record grid reference
for successful records

List successful records

Figure 2.6

given in Appendix A. Typical distribution maps are given in Figures 2.7 and 2.8. A range of plotting symbols is available.

2.6.7

ARCJWO17 MAPSTAFFS (MAP of STAFFordShire) Test Program (ALGOL 60)

This is a skeleton program to test the procedure MAPSTAFFS, a tailor-made routine which plots a map of Staffordshire complete with border, scale and grid north point. A distribution map may be plotted using specified National Grid References. Each reference is checked to be within the map area. The program is fully described in the Graphics chapter (paragraph 3.4.10).

2.6.8

ARCJWO19 Pottery profile suite (ALGOL 60)

This program calculates profile parameters, volumes and scale factors for wheel-made pots and records them on magnetic tape together with normalised coordinates for drawing the pots. This is the basis for an information retrieval scheme for pottery types. Recorded pots which are the most similar to any newly-found pot may be retrieved from the system. The comparisons are based on similarity measures, which are fully described in the Statistics and Pottery Classification chapters (see paragraphs 1.10.2 and 4.4.3). This method is to be used to analyse pottery from cave sites in Spain for Dr M.J.Walker.

2.6.9

ARCJWO21 Masons' marks analysis (ALGOL 60)

This program records scale factors and normalised coordinates for masons' marks on magnetic tape. Each significant end or intersection point of the geometric figures are identified using a specially-developed code. The number of ends and intersections, the coordinates of the centroid and the polar angles of the ends and intersections from the centroid are also recorded. This is the basis for an information retrieval scheme for masons' marks. Recorded masons' marks may be

ROMAN BRITAIN
SHOWING
CIVIL SETTLEMENTS,
LEGIONARY FORTRESSES,
NORTHERN FRONTIER WORKS
AND MAJOR FORTS

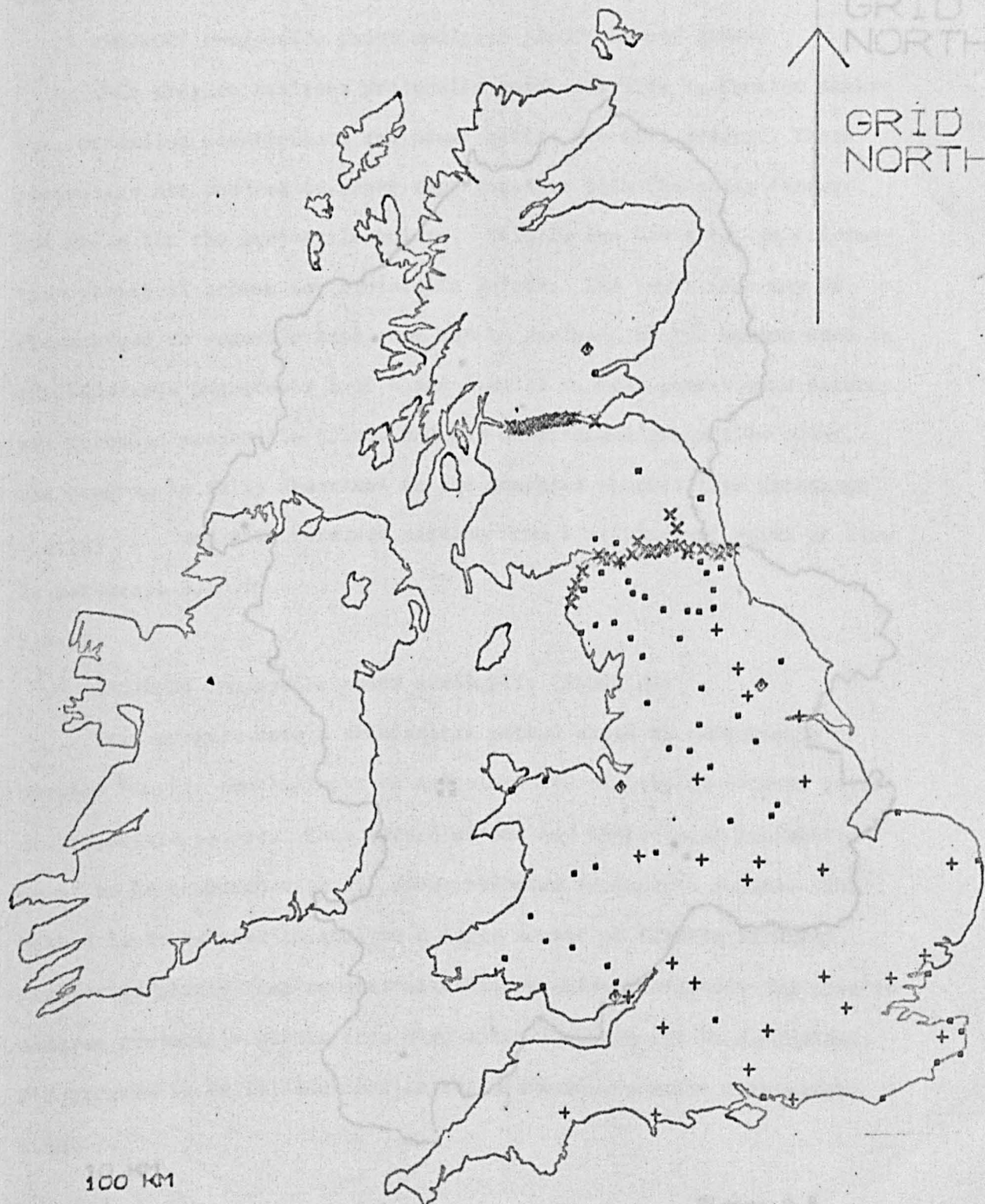
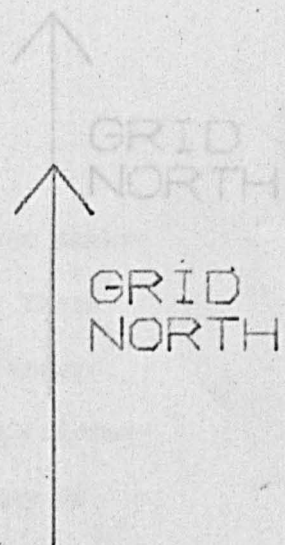


Figure 2.7

DISTRIBUTION OF ROMAN SITES IN THE COUNTY OF STAFFORDSHIRE

of ends and intersections and their locations relative to the centroids

The program is fully described in the Statistics chapter (paragraph

1.10.3).

2.6.10

ARCJWO27 Projectile point analysis (ALGOL 60 and EUGEN)

This program analyses projectile point profiles by Fourier analysis, producing coefficients and power series for the profiles. These

parameters are punched on paper tape together with the scale factors

and areas for the projectile points. This is the basis for an information

retrieval scheme for projectile points. The paper tape may be

transcribed to magnetic tape and may be analysed by the method used in

ARCJWO28 (see paragraphs 1.10.6 and 2.6.11) to make comparisons between

all recorded projectile points and any newly-found projectile point.

The program is fully described in the Graphics chapter (see paragraph

3.4.18) and receives mention from a statistical point of view

in paragraph 1.10.5.

2.6.11

ARCJWO28 Projectile point similarity (ALGOL 60)

This program uses a statistical method based on comparison

between Fourier coefficients to calculate the similarity between

all projectile points. This method allows any newly-found projectile

point to be compared with all other recorded projectile points. The

method is to be used to analyse a large number of Eastern European

projectile points from several sites for P. Allen-Jones, and also to

analyse projectile points from cave sites for Dr M.J. Walker.

The program is fully described in the Statistics chapter (paragraph

1.10.6).



GRID
NORTH

10 KM

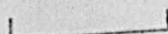


Figure 2.8

translated, rotated and scaled to give the best fit to any newly-found mason's mark. The comparisons are based on similarities in numbers of ends and intersections and their locations relative to the centroids. The program is fully described in the Statistics chapter (paragraph 1.10.3).

2.6.10

ARCJW027 Projectile point analysis (ALGOL 60 and EDGAR)

This program analyses projectile point profiles by Fourier analysis, producing coefficients and power series for five orders. These parameters are punched on paper tape together with the scale factors and areas for the projectile points. This is the basis for an information retrieval scheme for projectile points. The paper tape may be transcribed to magnetic tape, and may be analysed by the method used in ARCJW028 (see paragraphs 1.10.6 and 2.6.11) to make comparisons between all recorded projectile points and any newly-found projectile point. The program is fully described in the Graphics chapter (see paragraph 3.4.18) and also receives mention from a statistical point of view in paragraph 1.10.5.

2.6.11

ARCJW028 Projectile point similarity (ALGOL 60)

This program uses a statistical method based on comparisons between Fourier coefficients to calculate the similarity between pairs of projectile points. This method allows any newly-found projectile point to be compared with all other recorded projectile points. The method is to be used to analyse a large number of Eastern European projectile points from several sites for P.Allsworth-Jones, and also to analyse projectile points from cave sites in Spain for Dr M.J.Walker. The program is fully described in the Statistics chapter (paragraph 1.10.6).

2.6.12

SITIR1 Set up pottery check file (BASIC - designed for remote terminal use)

This is a program to write a check file for pottery of many archaeological periods to magnetic disc. The file is a hierarchic structure designed in conjunction with Mr P.Buckland of the University of Birmingham for use on a Roman/Medieval site in Doncaster. The system is to be developed for use by the Dept. of the Environment on the House of Commons excavation in London and on general excavations. The essence of the scheme is its multi-access character using a teletype in the site office connected to a remote computer. Very few archaeologists have used such a scheme in Britain.

If the file is to be written on a new disc or altered then the writing phase is entered. If not, SITIR4 is chained. In the writing phase the disc is erased and set in write mode. The read/write cycle is then entered. The numerical label of a classification group is read from the data list. If it is zero, the read marker is reset to the beginning of the file, the read mode is set up, and SITIR4 is chained. Otherwise the number of entries in the classification group is read, and this number of strings and next group labels are read from the list. The label, number of entries, and all pairs of strings and next group labels are written to disc. The read/write cycle then repeats.

2.6.13

SITIR4 Set up small finds check file (BASIC - designed for remote terminal use)

This is a program to write a check file for small finds to magnetic disc. The file is a hierarchic structure designed in conjunction with Mr P.Buckland of the University of Birmingham for use on a Roman/Medieval site in Doncaster.

Apart from the different data list and the use of a second disc file the program is identical to SITIR1. SITIR4 is chained from SITIR1, and itself chains SITREC.

2.6.14

*SITREC Site records analysis main program
(BASIC - designed for remote terminal use)*

This is a program to check and add records to a new or existing disc file for an archaeological site. There are also facilities for chaining SITIR2 or SITIR3.

By a modification of one statement before the program is run the site disc file may be erased and labelled with dummy identifiers acceptable to a later part of the program - this step is only necessary if the disc has been corrupted.

The normal entry to SITREC from SITIR4 immediately gives the option to enter SITIR2 or SITIR3 by chaining. If these options are not taken, the disc file read marker is set to the beginning of the file and the read mode is entered. Options are then given to make additions to the old file, create a new file (erasing the old file in the process) or terminate. The terminate option makes available the SITIR2 or SITIR3 alternatives again.

If the new file option is selected, the site identifier is read from the beginning of the old file. If it is a dummy identifier as created by the initialisation process above, no further action is taken. Otherwise the user is warned that he is attempting to delete an old file by the message:

FILE ALREADY ALLOCATED TO SITE < identifier >

He is then asked whether he still wishes to delete the file by the message:

DELETE YES/NO Y/N:?

If he answers Y, the password recorded on the old file is read, and

the user is asked to type in the old password. If the passwords tally, the file is erased. If the passwords do not tally, or if a N answer was given to the DELETE question, the program returns to the new file / old file / terminate option. When a password is accepted, the new site name and new password must be typed in, and these are written onto the disc. The program then enters the additions phase.

If the old file option is selected the site identifier is read from the disc. If it is a dummy identifier the message

FILE NOT ALLOCATED

is given and the program returns to the new file / old file / terminate option. Otherwise the message

FILE FOR SITE < identifier >

is given. The old password is read from the disc, and the user is asked to type in the old password. If the passwords tally the additions phase is entered, otherwise the program returns to the new file / old file / terminate option.

The additions phase commences by reading the first string from a record in the data list of SITREC. This should normally be a valid string, but if it is '0' the additions phase is terminated and the program returns to the new file / old file / terminate option. If a valid string is read, the sequential item number is printed. If the string is 'R' this means that the first six levels of the previous record are to be retained, and the next string goes into the seventh level. A string other than 'R' goes into the first level of the new record, and the next string goes into the second level.

If the next string is '*' this indicates the end of the record. Otherwise the string is checked to be one of the allowed entries for the classification in the required classification group stored by SITIR1 or SITIR4 on disc. The classification group consists of group

label, number of entries, and all pairs of strings and next group labels. As the string is checked, the next group label is taken from disc. Group labels always increase as the record is processed. If a next group label numerically less than the current group label is received, it is interpreted as an integer count, and this number of unchecked strings are read to complete the record. The records are checked to be the correct length and that all levels are acceptable (there are appropriate failure messages). When a record is successfully completed it is written to the disc and another is read, and this continues until '0' is read as the first level.

Typical records will now be illustrated:

<i>Present label</i>	<i>String</i>	<i>Check status</i>	<i>Level</i>	<i>Next label</i>
Pottery example				
10	'DON'	not checked	1	30
30	'ABC'	not checked ('-')	2	40
40	'STRATIFIED'	checked	3	50
50	'3'	not checked ('-')	4	60
60	'4'	not checked ('-')	5	70
70	'5'	not checked ('-')	6	75
75	'P'	checked	7	80
80	'ROMAN'	checked	8	270
270	'SAMIAN'	checked	9	280
280	'DEC.'	checked	10	5
	'DRAG.37'	not checked	11	4
	'23'	not checked	12	3
	'1'	not checked	13	2
	'2'	not checked	14	1
	'3'	not checked	15	0
	'*'	(next label checked to be 0)		

Small finds example				
10	'R'	replaced by previous level 1 ('DON') 75		
	'ABC'	retained	2	
	'STRATIFIED'	retained	3	
	'3'	retained	4	
	'4'	retained	5	
	'5'	retained	6	
75	'S'	checked	7	1010 (small find
1010	'COIN'	checked	8	1030 disc)
1030	'ROMAN'	checked	9	1040
1040	'AR'	checked	10	1050
1050	'DENARIUS'	checked	11	1
	'NERO'	not checked	12	0
	'*'	(next label checked to be 0)		

The full check lists generated by SITIR1 (pottery) and SITIR4 (small finds) on separate discs are given in full in Figures 2.9 and 2.10 respectively. Typical runs of SITREC are given in the program listing file.

2.6.15

SITIR2 Site records amendment (BASIC - designed for remote terminal use)

This is a program to amend records already written to disc by SITREC or NOCHEK.

The program asks the user to specify the number of an item he wishes to amend. The site identifier and password are read from the master disc and written to a work disc. Records are then read from the master disc and copied to the work disc until the required item is located.

If the master disc terminates before the item is found, the whole of the work disc is recopied to the master disc, the master disc read marker is set to the beginning of the disc and read mode is entered, while the work disc is erased and set to write mode. The program then returns to the stage where the user is asked to specify an item.

If the required item is found on the master disc it is printed out in full. Each level consists of the level number and contents. This is done so that the user may satisfy himself that he has located the correct item before he amends it.

Options are now given to alter, delete or reject the item. The reject option is used if an incorrect item has been located, or if no amendment is required. In this case the program returns to the stage where the user is asked to specify an item.

The delete option deletes the whole item from the master disc. This is done by omitting to write the located item to the work disc,

```

SITIR1
READY
LIST
10 REM SET UP CHECK FILE
20 FILES 1,2,3,4
22 PRINT "MODIFY CHECK FILE ON DISC 3 Y/N:"
24 INPUT D
26 IF D = "N" THEN 50
28 IF D <> "Y" THEN 22
30 SCRATCH ~3
32 READ A
33 PRINT A
34 IF A = 0 THEN 49
36 READ N
38 WRITE ~3, A, N
40 FURI = 1 TON
42 READ W, A
44 WRITE ~3, W, A
46 NEXT I
48 GOTO 32
49 RESTORE ~3
50 CHAIN SITIR4
1499 STOP
2030 DATA 30, 1, "-", 40
2040 DATA 40, 3, "STRATIFIED", 50, "UNSEALED", 50, "UNSTRATIFIED", 50
2050 DATA 50, 1, "-", 60
2060 DATA 60, 1, "-", 70
2070 DATA 70, 1, "-", 75
2075 DATA 75, 3, "P", 80, "S", 1010, "B", 1
2080 DATA 80, 5, "POST MED.", 90, "MED.", 180, "SAXON", 230, "ROMAN", 270
2081 DATA "PRE-ROMAN", 260
2090 DATA 90, 8, "CLAY PIPE", 100, "UTILITARIAN", 140, "STAFFS. SLIPWARE", 150
2091 DATA "WEDGWOOD", 0, "DELFT", 0, "BLUE ON WHITE", 170, "CREAMWARE", 0
2092 DATA "OTHER", 0
2100 DATA 100, 2, "STEM", 1, "BOWL", 1
2140 DATA 140, 3, "GREEN GLAZE", 3, "BROWN GLAZE", 3, "STONEWARE", 3
2150 DATA 150, 2, "PIPED DEC.", 160, "COMBED DEC.", 160
2160 DATA 160, 2, "BOWL", 3, "JUG", 3
2170 DATA 170, 2, "WILLOW PATTERN", 1, "OTHER", 1
2180 DATA 180, 8, "CISTERCIAN WARE", 190, "GREY", 190, "PINK", 190
2181 DATA "S. PENNINE GRITTY", 190, "BUFF", 190, "TUDOR GREEN", 190
2182 DATA "IMPORTED WARE", 210, "SHELLY GRIT WARE", 4
2190 DATA 190, 5, "SHERD", 200, "BASE", 2, "JUG", 2, "JAR RIM", 1
2191 DATA "OTHER", 3
2200 DATA 200, 3, "IMPRESSED DEC.", 1, "PLAIN", 1, "APPLIED STRIP DEC.", 1
2210 DATA 210, 3, "RAEREN STONEWARE JUG", 4, "POLYCHROME WARE", 4
2211 DATA "OTHER", 4
2230 DATA 230, 4, "PAGAN SAXON", 265, "MIDDLE SAXON", 265, "LATE SAXON"
2231 DATA 265, "SAXON-NORMAN", 250
2250 DATA 250, 2, "TORKSEY WARE", 3, "STAMFORD WARE", 3
2260 DATA 260, 5, "UNDIFF.", 265, "NEOLITHIC", 265, "N/BA", 265
2261 DATA "BRONZE AGE", 265, "IRON AGE", 265
2265 DATA 265, 2, "PLAIN", 3, "DEC.", 3
2270 DATA 270, 5, "SAMIAN", 280, "COLOUR COATED", 300, "MORTARIA", 350
2271 DATA "COARSE WARE", 380, "AMPHORA", 1
2280 DATA 280, 2, "PLAIN", 5, "DEC.", 5
2300 DATA 300, 2, "UNDIFF.", 4, "-", 310
2310 DATA 310, 4, "RHENISH", 320, "NENE VALLEY", 320, "SWANPOOL", 320
2311 DATA "OTHER", 320
2320 DATA 320, 8, "HUNT CUP", 4, "BASE", 3, "BEAKER RIM", 5
2321 DATA "BOWL RIM", 4, "FLAGON", 7, "JAR RIM", 6
2322 DATA "DEC. SHERD", 6, "-", 3
2350 DATA 350, 1, "-", 360
2360 DATA 360, 1, "-", 370
2370 DATA 370, 9, "UNTYPED", 2, "BEAD & ROLL", 2, "REEDED BEAD & ROLL", 2
2371 DATA "FLANGED", 2, "REEDED FLANGE", 2, "HAMMERHEAD", 2
2372 DATA "REEDED HAMMERHEAD", 2, "WALL-SIDED", 2, "-", 1

```

Figure 2.9

Sheet 1 of 2

Check file for pottery

2380DATA380,2,'CALCITE GRITTED WARE',390,'-',430
 2390DATA390,4,'SHERD',400,'BASE',1,'DALES WARE',1,'-',410
 2400DATA400,2,'PLAIN',1,'DEC.',1
 2410DATA410,8,'JAR RIM',420,'SWANPOOL TYPE H',1
 2411DATA'PLAIN RIMMED DISH',1,'BOWL',1,'FLANGED BOWL',1
 2412DATA'SIGNAL STATIONS WARE',1,'HUNTCLIFF JAR',1,'-',0
 2420DATA420,5,'EVERTED',1,'CLUB',1,'CAVETTO',1,'RECURVED',1
 2421DATA'LID SEATED',1
 2430DATA430,6,'BLACK BURNISHED WARE',440,'GREY BURNISHED WARE'
 2431DATA440,'GREY WARE',440,'WHITE FABRIC',440,'BUFF FABRIC',440
 2432DATA'OXIDISED',440
 2440DATA440,5,'JAR',460,'BEAKER',450,'BOWL',530,'DISH',530,'-',3
 2450DATA450,2,'HANDLE',1,'-',1
 2460DATA460,4,'RIM',470,'BASE',510,'SHERD',520,'HANDLE',1
 2470DATA470,5,'CAVETTO',480,'RECURVED',480,'SPLAYED',480
 2471DATA'LID SEATED',490,'-',590
 2480DATA480,2,'OFFSET',590,'PLAIN',590
 2490DATA490,4,'DERBYSHIRE WARE',590,'SWANPOOL TYPE H',590
 2491DATA'S. YORKS',590,'OTHER',590
 2510DATA510,4,'PLAIN',2,'PEDESTAL',2,'RING-PEDESTAL',2
 2511DATA'RING BASE',2
 2520DATA520,5,'PLAIN',2,'RILLED',2,'LATTICE DEC.',2
 2521DATA'RUSTIC',2,'OTHER',2
 2530DATA530,3,'RIM',540,'SHERD',570,'BASE',580
 2540DATA540,4,'CULANDER',550,'FLAT',560,'FLAT & REEDED',560
 2541DATA'-',550
 2550DATA550,2,'FLAT RIM',590,'ROUNDED BASE',590
 2560DATA560,3,'CARINATED',590,'NOT CARINATED',590,'UNKNOWN',590
 2570DATA570,3,'PLAIN',590,'LATTICE DEC.',590,'OTHER DEC.',590
 2580DATA580,5,'KICKED',590,'FLAT CHAMFERED',590,'PEDESTAL',590
 2581DATA'RING',590,'FOOTRING',590
 2582DATA590,9,'BBW1',2,'BBW2',2,'S. YORKS',2,'LINCS',2
 2583DATA'E. YORKS',2,'MIDLAND',2,'S.E. ENGLAND',2,'DERBY',2
 2584DATA'OTHER',2
 2590DATA0
 9999END

Figure 2.9

Sheet 2 of 2

Check file for pottery

1010DATA1010,9,'BUNE',1020,'COIN',1030,'GLASS',1120,'BRONZE',1140
 1011DATA'SILVER',1190,'STONE',1220,'POTTERY',1270,'GOLD',1300
 1012DATA'IRON',1310
 1020DATA1020,5,'PIN',0,'COMB',0,'COUNTER',0,'HANDLE',0,'DICE',0
 1030DATA1030,8,'ROMAN',1040,'SAXON',0,'NORMAN',0,'ANJEVIN',0
 1031DATA'MED.',1110,'POST MED.',0,'TUDOR',0,'JETTON',0
 1040DATA1040,4,'AR',1050,'AE',1060,'AE(SILVERED)',1080,'AU',1100
 1050DATA1050,2,'DENARIUS',1,'SILIGUA',1
 1060DATA1060,2,'SEMIS/B. TETRADRACHM.',1,'LARGE AE',1070
 1070DATA1070,3,'SESTERTIUS',1,'DUPONDIIUS',1,'AS',1
 1080DATA1080,2,'BARB. RADIATE',0,'SMALL AE',1090
 1090DATA1090,4,'ANTONINIANUS(RADIATE)',1,'FULLIS',1,'AE3',1,'AE4',1
 1100DATA1100,2,'AUREUS',1,'SOLIDUS',1
 1110DATA1110,4,'2D',0,'1D',1,'1/2D',1,'1/4D',1
 1120DATA1120,2,'RINGSTONE',0,'BEAD',1130
 1130DATA1130,2,'MELON',0,'OTHER',0
 1140DATA1140,12,'MILITARY EQUIPMENT',1150,'FRAG. OBJECT',0
 1141DATA'BRACELET',1160,'SPOON',0,'VESSEL',0,'EARPROBE',0
 1142DATA'MANICURE SET',0,'TWEEZERS',0,'NAILCLEANER',0
 1143DATA'PIN',0,'ROMAN',1200,'MED.',1210
 1150DATA1150,12,'SHIELD EDGING',0,'HARNESS FITTING',0,'STRAP END',0
 1151DATA'STRAP DISTRIBUTOR',0,'PENDANT',1160,'BIT',0
 1152DATA'PHALERA',0,'HINGE',0,'BUCKLE',1170,'STUD',1160,'CHAPE',1180
 1153DATA'OTHER',0
 1160DATA1160,2,'PLAIN',0,'DEC.',0
 1170DATA1170,2,'SIMPLE',0,'ZOOMORPHIC',0
 1180DATA1180,2,'SWORD',0,'DAGGER',0
 1190DATA1190,2,'ROMAN',1200,'MED.',1210
 1200DATA1200,3,'BROOCH',1,'LOCK FRAGMENT',0,'RING',0
 1210DATA1210,3,'RING',0,'BUCKLE',0,'HARNESS TRAPPING',0
 1220DATA1220,6,'JET',1230,'SPINDLE WHORL',0,'WHETSTONE',1260
 1221DATA'MOULDDING',1260,'WORKED STONE',1260,'QUERN',1240
 1230DATA1230,4,'FRAG.',0,'BEAD',0,'PIN',0,'BRACELET',0
 1240DATA1240,2,'FRAG.',1250,'COMPLETE',1250
 1250DATA1250,2,'LAVA',1260,'MILLSTONE GRIT',1260
 1260DATA1260,3,'RB',0,'MED.',0,'POST MED.',0
 1270DATA1270,7,'STAMPED CLAY PIPE',0,'POST MED. POTTERY STAMP',0
 1271DATA'STAMPED',1280,'GRAFFITI',1290,'SPINDLE WHORL',1290
 1272DATA'COUNTER',1290,'OTHER',0
 1280DATA1280,3,'SAMIAN',1,'MORTARIUM',0,'--',0
 1290DATA1290,2,'SAMIAN',0,'OTHER',0
 1300DATA1300,3,'RING',0,'GLOBULE',0,'OBJECT',0
 1310DATA1310,2,'MED.',1320,'ROMAN',1320
 1320DATA1320,10,'OBJECT',0,'LYNCH PIN',0,'RIVET',0,'ARROWHEAD',0
 1321DATA'KNIFE',0,'SPEARHEAD',1330,'BUCKLE',0,'SHIELD BOSS',0
 1322DATA'PLUMBATUM',0,'OTHER',0
 1330DATA1330,3,'PILUM',0,'HASTA',0,'OTHER',0
 1340DATA0
 9999END

Figure 2.10

Check file for small finds

but by continuing to read subsequent items from the master disc and to write them to the work disc. When the master disc terminates the recopying procedure is carried out as above.

The alter option asks the user to specify the number of a level to be amended. He is then asked to type the new contents of the level, which are then entered into the record without checking. The program continues to ask for further levels to be amended until level 0 is specified. This causes the amended item to be written to the work disc. Subsequent items are read from the master disc and written to the work disc until the master disc terminates, when the recopying procedure is carried out (see above).

After each alteration, deletion or rejection the program continues to ask for further items to be located until item 0 is specified. This gives the options to chain SITIR3 and to chain SITREC or NOCHEK as applicable. Other options are to continue with SITIR2 or to terminate.

Typical runs of SITIR2 are given in the program listing file.

2.6.16

SITIR3 Site records listing and retrieval (BASIC - designed for remote terminal use)

This is a program to list records and to control information retrieval runs on records already written to disc by SITREC or NOCHEK.

The program reads the site identifier and password from the master disc, then gives the options to list, carry out an information retrieval run, or to exit from SITIR3.

The list option asks the user to specify the number of the item at which the listing is to start. Records are read from the master disc until the item is located.

If the master disc terminates before the item is found, the master disc read marker is set to the beginning of the disc, read mode is

entered, the site identifier and password are read and the program returns to the list/IR/exit options.

If the required item is found on the master disc, this item and all subsequent items are listed, each identified by the item number. The program returns to the list/IR/exit options.

The information retrieval option examines every item on the master disc to see if it satisfies a logic check. If it does, the item is printed, otherwise the next item is checked. When all items have been checked, the program returns to the list/IR/exit options. The logic check is designed by the user and inserted using the program alteration facility of the BASIC system. It is usual to design a different logical check for each run; the extra statements are deleted at the end of the run and do not modify the program on the systems disc unless the command SAVE is used. Comments define the slot in the program into which these extra statements are to be inserted (line numbers 640-698, see listing). A typical logical check is given below:

```
640 IF L$(3) ≠ 'STRATIFIED' THEN 800 (800 is the failure exit)
642 IF L$(11) = 'DRAG.37' THEN 700 (700 is the success exit)
644 IF L$(11) ≠ 'DRAG.29' THEN 800 (if = then goes by default to
                                700, the next sequential statement)
```

Where L(n)$ is the nth level string variable.

This check locates all items which have *stratified* pottery of *Drag.37* or *Drag.29* forms, or in Boolean algebraic terms

$$f = \text{Stratified} \wedge (\text{Drag.37} \vee \text{Drag.29})$$

The exit option gives the options to chain SITIR2 and to chain SITREC or NOCHEK as applicable. Other options are to continue with SITIR3 or to terminate while still holding the SITIR3 program in core, allowing the modifiable Boolean check to be inserted. When this has

been done, SITIR3 may be run again by typing the command RUN.

Typical runs of SITIR3 are given in the program listing file.

2.6.17

*NOCHEK Site records analysis main program without checking
(BASIC - designed for remote terminal use)*

This is a program to add records to an existing disc file for an archaeological site. There is no check on the records as for SITREC - it has been found that the checking process consumes a lot of computer time on the multi-access system.

By a modification of one statement before the program is run the site disc file may be erased and labelled with dummy identifiers acceptable to a later part of the program - this step is necessary when opening a new file or when the file has been corrupted.

The normal entry immediately gives the option to enter SITIR2 or SITIR3 by chaining. If these options are not taken, the disc file read marker is set to the beginning of the file and the read mode is entered. The site identifier and password are read from the disc. The message

FILE FOR SITE < identifier >

is printed. The user is asked to type in the password. If the passwords tally the additions phase is entered, otherwise the program returns to the normal entry.

The additions phase commences by reading the first string from a record in the data list of NOCHEK. This should normally be a valid string, but if it is '0' the additions phase is terminated and the program returns to the normal entry. If a valid string is read, the sequential item number is printed. If the string is 'R' this means that the first six levels of the previous record are to be retained, and the next string goes into the seventh level. A string other than

'R' goes into the first level of the new record, and the next string goes into the second level.

If the next string is '*' this indicates the end of the record. Otherwise another string is read. When a record is complete it is written to the disc and another is read, and this continues until 'O' is read as the first level.

Where elaborate checking may be dispensed with this program efficiently replaces SITREC as the main program of the remote terminal site records suite. Typical runs of NOCHEK are given in the program listings file.

Information Retrieval Bibliography

General

- Anson-Jonker *Anson-Jonker equipment for information retrieval
Geo-stat technique*
George Anson & Co.Ltd, London
- Atkinson,D. and Oswald,A. London clay tobacco pipes
Journal of the British Archaeological Association
171-227
- Bergengren,G., 1970/71 Automatic data processing in the registration
of museum collections in Sweden
(UNESCO) Museum 23(1), 56-58
- Bordaz,J. and Bordaz,V., 1965/66 A critical examination of data processing in
archaeology, with an evaluation of a new
inverted data system
American Antiquity 31, 494-501
- Chenhall,R.G., 1966 S37. Museum cataloguing by computer
C.Hum.
in *Directory of Scholars Active*, 1966/67, 240
- Chenhall,R.G., 1967 The description of archaeological data in
computer language
American Antiquity 32(2), 161-167
- Chenhall,R.G., 1968 The analysis of museum systems
in *The Metropolitan Museum Conference on
Computers and their Potential Applications
in Museums*
New York, April 1968
- Chenhall,R.G., 1970 S100. The Arkansas archaeological data bank
C.Hum.
in *Directory of Scholars Active*, 1970, 351
- Chenhall,R.G., 1971a Note in
Newsletter of Computer Archaeology 6(4), 1
- Chenhall,R.G., 1971b The archaeological data bank: a progress report
C.Hum. 5(3), 159-169
- Chouraqui,E., 1972 IR1, Archéologie Romaine
C.Hum. 6(3)
in *Humanities Programs Available*, 181
- Cuisenier,J., 1970/71 Feasibility of using a data processing system in
the Musée des Arts et Traditions Populaires,
Paris
(UNESCO) Museum 23(1), 33-36

- Ellin,E., 1968a V10. Museum computer network
C.Hum. 2(5)
in *Directory of Scholars Active 1967/68*(2), 250
- Ellin,E., 1968b V10. Museum computer network
C.Hum.
in *Directory of Scholars Active 1968/69*(1), 319
- Ellin,E., 1968c An international survey of museum computer activity
C.Hum. 3(2), 65-86
- Ellin,E., 1968d Information systems and the humanities: a new renaissance
in *The Metropolitan Museum Conference on Computers and their Potential Applications in Museums*
New York, April 1968
- Ellin,E., 1968e *Report of the Museum Computer Network Project*
New York
- Ellin,E., 1969 V10. Museum computer network
C.Hum.
in *Directory of Scholars Active 1968/69*(2), 141
- Gaines,S.W., 1971a Note in
Newsletter of Computer Archaeology 6(4), 1
- Gaines,S.W., 1971b Computer application in an archaeological field situation
Newsletter of Computer Archaeology 7(1), 2-4
also reprinted in
MASCA Newsletter 7(2), 2-3
- Gardin,J.C., 1965 Reconstructing an economic network in the ancient East with the aid of a computer
in Hymes,D. (ed), 1965
The use of computers in anthropology
377-391, discussion 393-397
Mouton & Co., The Hague 558 pp.
- Green,D.F., 1966 S30. Development of a cataloguing-research system for the University of Missouri, Anthropology Museum
C.Hum.
in *Directory of Scholars Active 1966/67*, 238
- Gundlach,R., 1966 1966 ARDOC: A system of information storage and retrieval in archaeology
Proceedings of the international symposium on mathematical and computational methods in the social sciences
International Computation Center, Rome
- Hebditch,D.L., 1972 A simple language for a data-base
Computer Bulletin 16(1), 36-40

- Hulton,F.C., 1968 PEEKABIT, Computer offspring of punched card
PEEKABOO, for natural language searching
Comm. A.C.M. 11(9), 595-598
- IRGMA Newsletter*, 1969 1, January 1969
Information Retrieval Group of the Museums
Association, London
- Jonker,F., 1963 The new Termatrix line of I.R. Systems - the
Minimatrex line of I.R. systems
American Documentation 14(4), 276-282
- Laflin,S. (Mrs P.Barker), 1972 *Archaeological Gazetteer for Shropshire*
University of Birmingham
- Lavell,C., 1967 The Council for British Archaeology and
information retrieval
Museums Journal 67, 113-114
- Lewis,G.D., 1965 Obtaining information from museum collections
and thoughts on a national museum index
Museums Journal 65, 12-22
- Lewis,G.D., 1970/71 An interdisciplinary communication format
for museums in the United Kingdom
(UNESCO) *Museum* 23(1), 24-26
- Lewis,G.D., et al., 1967 Information retrieval for museums
Museums Journal 67, 88-120
- Lewis,G.D., et al., 1969 *Draft proposals for an interdisciplinary
museum cataloguing system*
Information Retrieval Group of the Museums
Association London, 23rd April 1969, 46 pp
- Morrison,D.R., 1968 PATRICIA - Practical Algorithm To Retrieve
Information Coded In Alphanumeric
Journal A.C.M. 15(4), 514-534
- Newman,W., 1969 Correspondence with Chenhall,R.G. in
Newsletter of Computer Archaeology 5(2), 1
- O'Connor,D., Voigt,M.M. and Wyuker,E., 1968 Information retrieval for archaeological sites
MASCA Newsletter 4(1), 2-3
- Perring,F.H., 1967 The requirements of the research worker in
biology
Museums Journal 67, 108-111
- Renfrew,C., 1967 The requirements of the research worker in
archaeology
Museums Journal 67, 111-113
- Rensberger,J.M. and Berry,W.B.N., 1967 An automated system for the retrieval of
museum data
Curator 10(4), 297-317
American Museum of Natural History

- Reynolds,C.F., 1970 Meeting on CODIL
Advanced programming study group
Computer Bulletin 14(7), 244-245
- Reynolds,C.F., 1971a CODIL, Part 1. The importance of flexibility
Computer Journal 14(3), 217-220
- Reynolds,C.F., 1971b CODIL, Part 2, The CODIL language and its
interpreter
Computer Journal 14(4), 327-332
- Reynolds,C.F., 1971c Handling cave fauna records on a computer
Trans. Cave Research Group of Great Britain
13(3), 160-165
- Roads,C.H., 1968 Data recording, retrieval and presentation
in the Imperial War Museum
Museums Journal 68, 277-283
- Rusco,M., 1968 Index for the Nevada archaeological survey
archives
MASCA Newsletter 4(1), 2
- Schneider,M.J., 1968 S30. Catalogue-research system for the
University of Missouri Museum of Anthropology
C.Hum. 2(5)
in *Directory of Scholars Active 1967/68*(2), 247
- Schneider,M.J., 1971 Archaeological data banks
Letter to the Editor
C.Hum. 5(4), 239-241
- Sokal,R.R. and Sneath,P.H.A., 1966
Efficiency in taxonomy
Taxon 15, 1-21
- Soper,J.H. and Perring,F.H., 1967
Data processing in the herbarium and museum
Taxon 16(1), 13-19
- Vance,D., 1970 Museum data banks
Inform. Stor. Retr. 5, 203-211
Pergamon Press
- Vance,D., 1971 Museum Computer network
Letter to the Editor
C.Hum. 6(1), 47-48
- Vance,D. and Heller,J., 1971
Structure and content of a museum data bank
C.Hum. 6(2), 67-84
- Walker,M.J., 1970a An analysis of British petroglyphs
Science and Archaeology 2/3, 30-61
reviewed by Scollar I., 1972
C.Hum. 6(3), 191
- Walker,M.J., 1970b An analysis of British petroglyphs
Note in
Science and Archaeology 4, 26-29 N4/1

Wilcock, J.D., 1970a

Information retrieval for cave records
Trans. Cave Research Group of Great Britain
12(2), 96-98
also printed in
Science and Archaeology 6, 9-11

Wilcock, J.D., 1970b

Petroglyphs by computer
Science and Archaeology 2/3, 27-29
reviewed by Scollar, I., 1972
C.Hum. 6(3), 191

56 references

Information Retrieval Bibliography

Roman Inscriptions

- Chouraqui, E. et al., 1972 L477. Automatic processing of the Corpus of Latin Inscriptions
C.Hum.
in Directory of Scholars Active
- Collingwood, R.G., 1921-1938 Roman Britain in 19.., II. Inscriptions
Journal of Roman Studies 11
12
14-19
21-28
- Collingwood, R.G. and Wright, R.P., 1965
The Roman Inscriptions of Britain
I. Inscriptions on stone
Oxford 790 pp
- Conference on Roman epigraphy, 1967
Punching of cards: I
Data concerning the epitaph and the grave
- Dessau, H. (ed.), 1892-1916 Inscriptiones Latinae Selectae
3 Volumes in 5 parts
- Haverfield, F.J., 1892 Additamenta quarta ad CIL vii
Ephemeris Epigraphica 7, 273-354
- Haverfield, F.J., 1913a Additamenta quinta ad CIL vii
Ephemeris Epigraphica 9, 509-690
- Haverfield, F.J., 1913b Roman Britain in 1913
British Academy Supplemental paper 2
Oxford
- Haverfield, F.J., 1914 Roman Britain in 1914
British Academy Supplemental paper 3
Oxford
- Hübner, E., 1873 Inscriptiones Britanniae Latinae
Corpus Inscriptionum Latinarum (CIL) vii
Berlin
- Hübner, E., 1877a Additamenta prima ad CIL vii
Ephemeris Epigraphica 3, 113-155
- Hübner, E., 1877b Additamenta secunda ad CIL vii
Ephemeris Epigraphica 3, 311-318
- Hübner, E., 1881 Additamenta tertia ad CIL vii
Ephemeris Epigraphica 4, 194-212

- Jory,E.J., 1968 S83. An index to Vol.VI of the Corpus
Inscriptionum Latinarum
C.Hum.
in Directory of Scholars Active 1968/69(2), 318
- Jory,E.J. and Moore,D.W., 1966
An index to CIL VI: Two aspects
Revue 2, 7-16
Organisation Internationale pour l'Etude
des Langues Anciennes par Ordinateur
- Oehler,H., 1970/71 Electronic documentation of a collection of
Roman sculpture photographs
(UNESCO) Museum 23(1), 46-52
- Rivet,A.L.F., 1965 *Notes on Roman Epigraphy*
Department of Classics, University of Keele
- Stefan,A., 1971 Applications of mathematical models to
epigraphy in Hodson,F.R., Kendall,D.G. and Tautu,
P. (eds), 1971
*Mathematics in the archaeological and historical
sciences*
Edinburgh University Press, Edinburgh 267-275
- Thompson,F.H., 1968 The zoomorphic pelta in Romano-British art
Antiquaries Journal 48,47-58
- Wright,R.P., 1939-1970 Roman Britain in 19.., II. Inscriptions
Journal of Roman Studies 29-60

20 references

Information Retrieval Bibliography

Ribchester Museum Catalogue

- Abram, W.A., 1876a Account of a Roman sculptured slab found at Ribchester
Trans. Hist. Soc. Lancashire and Cheshire
3rd series, 4, 190-193 and plate
(RIB 595 Horseman and recumbent foe)
- Abram, W.A., 1876b A Roman sculptured slab found at Ribchester
Proc. Soc. Antiq. 2nd series, 7, 30-34
(RIB 595 Horseman and recumbent foe)
- Andrew, W.J., 1903 Buried treasure, some traditions, records and facts
J. Brit. Arch. Assoc. 59, 8-32
reprinted in
Brit. Num. J. 1, 9-59 and 4 plates
(Ribchester helmet, now in British Museum)
- Anon., 1840 Gold coin of Trajan, found at Ribchester
Numismatic Chronicle 3, 60
- Anon. Objects found at Ribchester not now preserved in the museum
(26 stone objects, 13 metal objects, 9 miscellaneous objects)
- Atkinson, D., 1929 The southern corner of the fort at Ribchester
Journal of Roman Studies 19, 191-192
- Axon, W.E.A., 1907 Apollo Maponus, Mabon and the legend of the oldest creatures
Trans. Lancashire and Cheshire Ant. Soc.
24, 105-112
(RIB 583 altar to Apollo Maponus)
- Bell, W., 1852 On the bronze mask attached to the Roman helmet found at Ribchester
Gentlemen's Magazine N.S. 38, 450-453
(Ribchester helmet, now in British Museum)
- Collingwood, R.G. and Wright, R.P., 1965
The Roman Inscriptions of Britain (RIB)
I. Inscriptions on stone 194-201
Oxford 790 pp.
(RIB 583-599, 2268, 2269)
- Conway, R.S., 1908 Excavation at Ribchester
Classical Review 22, 196-197
- Edwards, B.J.N., Jones, G.D.B. and Webster, P.V., 1969
Ribchester Playing Field Excavation, 1969
Ribchester Museum
(Vicus area of timber buildings, later a parade ground for the Sarmatian cavalry)

- Edwards, B.J.N., Jones, G.D.B., Webster, P.V. and Wild, J.P., 1970
Excavations at Ribchester Fort 1970
on behalf of the Ministry of Public Buildings
and Works and Manchester University
 Ribchester Museum
 (Praetorium, later barracks, west gateway
 and defences)
- Garstang, J., 1899a
Report of excavations on the site of Roman
Ribchester (BREMETENNACVM)
 Ribchester Excavation Fund, Preston
- Garstang, J., 1899b
 Roman Ribchester
Antiquary 35, 80-84
- Garstang, J., 1900
 The Ribchester excavations
Antiquary 36, 184-186
- Garstang, J., 1902a
Roman Ribchester: Excavations, 1898
Interim Report
 Ribchester Excavation Fund, Preston
 (Granaries and adjoining buildings. West gate)
- Garstang, J., 1902b
 List of Roman remains from Ribchester
Trans. Hist. Soc. Lancashire and Cheshire
 53, 189-216 and 2 plans
 (The Ribchester "temple")
- Haverfield, F.J., 1904
 The Ribchester "temple"
Trans. Hist. Soc. Lancashire and Cheshire
 54, 197-202
- Heathcote, W.H., 1890
 Ribchester
Trans. Lancashire and Cheshire Antiq. Soc.
 7, 229-235 and plate
- Hinde, J.H., 1855
 BREMETENNACVM: On the site of Bremetennacum
 in the Notitia and Bretonacum in the
 Itinerary, and on the bearing of this question
 on Horsley's theory of secondary stations
 "per lineam valli", with some speculations as
 to the sites of certain other stations in the
 Notitia and Itinerary
Archaeologia Aeliana 1st series, 4, 109-118
- Hopkinson, J.H., 1928
The Roman fort at Ribchester
 Third edition, revised and enlarged by Atkinson,
 D. Published for the excavation committee of the
 Manchester and district branch of the Classical
 Association
 Manchester University Press, Manchester 35 pp.
 (list of inscriptions)

- Jones, G.D.B., 1970 *Ribchester Roman Fort*
(Known features of the fort up to the excavations of 1969)
- Just, J. and Harland, J., 1850 On Roman Ribchester
J. Brit. Arch. Ass. 6, 229-251
- May, T., 1908 Excavations at Ribchester
Trans. Lancashire and Cheshire Ant. Soc. 25, 211-217
(Principia on the site of Churchgates, excavated during the conversion of the Church Crofts property for Mr Richard Greenall and Miss Margaret Greenall, founder of the Ribchester museum)
- National Trust, 1966 *The Roman Fort at Ribchester*
Country Life Ltd, London
for the National Trust 29 pp.
- Richmond, I.A., 1943 Maponus
Archaeologia Aeliana 4th series, 21, 206-210
(Apollo-Maponus, the god worshipped at Ribchester, see RIB 583)
- Richmond, I.A., 1945 The Sarmatae, Bremetennacum Veteranorum and the Regio Bremetennacensis
Journal of Roman Studies 35, 15-29
(Discussion of the origin of the Sarmatian cavalry (cf. the Thracian horseman) from Romania and the Transylvanian plain. The administration of the region from Ribchester, which was possibly also the religious centre of the region. Horse-raising in the Fylde region, originally fenlands).
- Richmond, I.A., 1955 *Roman Britain*
Pelican Books A315 240 pp.
(reference to Ribchester, p.54 and to Maponus, p.196)
- Sidebotham, S., post 1925 *Bremetennacum*
A story of Ribchester during the Roman occupation from the year 80 A.D. to 380 A.D.
Ribchester Museum
- Smith, T.C. and Shortt, J., 1890 *The history of the Parish of Ribchester*
London 1-41 and 270
- Taylor, M.V. and Collingwood, R.G., 1927 Roman Britain in 1927
Journal of Roman Studies 17, 184-219 and plate
- Towneley, C., 1815 Figures of an antique helmet and mask of bronze and other antiquities discovered at Ribchester in Lancashire
Vetusta Monumenta 4, 1-2 and plates 1-4
Society of Antiquaries, London

Watkin,W.T., 1878

Roman Ribchester
Trans. Hist. Soc. Lancashire and Cheshire
3rd series, 6, 1-26 and plate

Watkin,W.T., 1883

Roman Lancashire
Liverpool 125-163

Weston,S., 1798

Observations on Mr Towneley's antique
bronze helmet
Archaeologia 13, 223-226

35 references

Chapter 3

GRAPHICS IN ARCHAEOLOGY

3. *Graphics in Archaeology*

3.1

The archaeologist continually has a need to produce diagrams for his publications. Typical diagrams are drawings of finds, pottery profiles, site plans, sections, distribution maps and statistical aids such as histograms, piecharts, dendrograms, scalograms and tables. To be made presentable for publication these need such features as borders, scales, north points and legends in a variety of type faces; the need to provide even such mundane features occupies many hours of the archaeologist's time which could be more profitably spent. It is therefore remarkable that computer graphics has not been more widely used by the archaeologist for the production of diagrams. This thesis develops a graphics system for archaeologists, and a large number of the Figures have been produced by computer using this system.

3.2

Previous graphics applications, mainly developed for purposes other than archaeology, will be described in section 3.3. Graphics programs developed for this thesis which control the digital incremental plotter and CRT display unit will be described in section 3.4.

3.3 *Previous graphics applications*

3.3.1

The SKETCHPAD man-machine graphical communication system was developed in the U.S.A. in 1963. The developments are described in three papers (Sutherland, 1963; Johnson, 1963; Denil, 1966). I.E. Sutherland describes the prototype SKETCHPAD system which eliminates typed statements, except for the specification of legends. Lines, circles and arcs may be generated. *Subpictures* may be generated and

stored for future use; on retrieval they may be duplicated in many different areas of the display. A *ring structure* is used to link the features of geometrical constructions, e.g. the end points of a line segment may be found by following pointers in the line block, and all line segments which terminate at a given point may be found by following pointers in the point block (the list closes on itself, hence the term *ring structure*). Deletion of items releases storage in the *display file*, which is later tidied up by a *garbage collector* routine. A *light pen*, which is a flexible light-tube connected to a photosensitive device inside the display console, can be used to sense illuminated features on the display screen, and thus identify parts of the diagram for deletion or modification, or specify new points (see Figure 3.1). Since the light pen can only detect light already produced by the display, however, new points can only be specified by arranging for a tracking routine to continually generate and display a *tracking cross* at the last detected position of the light pen. If the operator moves the light pen too quickly for the tracking routine to respond, a *spiral raster* is generated starting at the last known position of the pen, and if this fails to locate the pen a *TV raster* is generated which scans the whole screen and must find the pen if it is still being held against the screen. The light pen may also be used to select items from a *menu* or one of a series of *light buttons* which are groups of characters specifying particular routines (these change the path of the software) or reduced subpictures. Light buttons have been used in several of the graphics programs developed for this thesis (see paragraphs 3.4.15, 3.4.16 and 3.4.17). One use of the light pen is to specify the positions of legends. SKETCHPAD generates legends of any character size and orientation. Constraints on particular items may be displayed as lines linking the constrained items.

Line-drawing
display unit

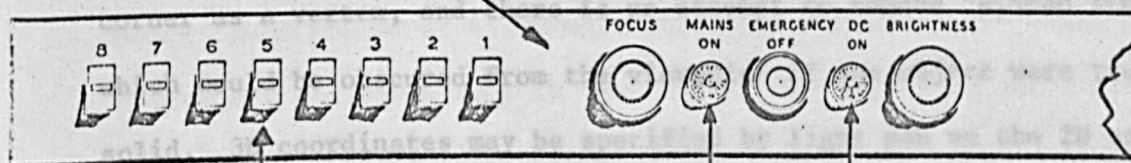
Screen

Light pen

Device
control unit
in
pedestal

Control
panel

Light pen
activating
foot switch



Function switches

Indicator lamps

Recursive expansion, merging and deletion are available; e.g. this means that subpictures may be generated within subpictures. When geometric features are being drawn, the constraints may be specified first, and then the sketched figure is fitted to them. One feature which saves a lot of expense is the generation of repetitive patterns of complex design. The computer can display subpictures of any complexity in a repetitive sequence almost instantaneously, whereas to have these produced by conventional drawing-office techniques is very costly indeed. Moving pictures may be generated by repetitive recalculation and display, and thus figures may be made to rotate. Sutherland's general conclusions on the use of computer graphics for diagrams are that it is only worthwhile if extra information is generated at the same time, e.g. design data. In archaeological terms this means that statistics should be calculated at the same time as diagrams are drawn, and this has been done in the projectile point analysis program (see paragraph 3.4.18), where Fourier analysis is carried out at the same time as the profiles are being drawn.

Johnson extends SKETCHPAD to handle coordinates in three dimensions, and it is then renamed SKETCHPAD III. 3D objects are displayed in four views in the four quadrants of the screen, viz. perspective view, top view, front view and side view. The light pen is used to delete and move items. There are expansion, contraction and rotation features as well as a facility to change the observers viewpoint for the perspective display (*tracking*). The "wireframe" method of representing solid objects is used, i.e. each edge appears as a line, each corner as a vertex, and there is no attempt to remove "hidden lines", which would be obscured from the viewpoint if the object were truly solid. 3D coordinates may be specified by light pen on the 2D screen by using the rotation facility and drawing new lines always in the

plane of the screen. A point may be positioned on an existing line identified by light pen but not lying in the plane of the screen, by associating a sphere (say of 0.3 cm radius) with the tracking cross centre. The light pen is sensitive within a circle of this radius. When the line intersects the sphere, the point is identified as that point on the line which is nearest the tracking cross centre, 2D coordinates are assigned in the plane of the screen, and the coordinate in the third dimension is calculated from the equation of the line. Thereafter, a new line starting at this point will be specified as lying in a plane parallel to the plane of the screen, and will retain the depth coordinate of the point identified on the first line (unless the line were to be terminated at a point on another prespecified line). Alternatively, a cylinder with axis perpendicular to the screen plane may be used, and the point is taken as the point on the line nearest to the axis of the cylinder at the intersection with the surface of the cylinder. Thus the depth coordinate need not be in the plane of the screen. In the case where lines are drawn superimposed on the 2D screen (but not necessarily intersecting in 3D) the line identified by the light pen will be indeterminate. Although the lines appear to the eye to be illuminated together, this is an optical illusion caused by the relatively slow response (0.04 sec) of the eye and the medium persistence phosphor with an optical peak often in the orange or green wavelengths. The light pen is sensitive in the ultra-violet and views an entirely different phosphor of very short persistence. Thus the two superimposed lines are never illuminated together as far as the light pen is concerned, and which line is seen first will depend on the timing of the refresh display file scan when the light pen becomes active and the position of the lines in the display file. The depth coordinate

for the new line is taken from the *first* line seen. SKETCHPAD III has a facility to store the current display file on magnetic tape for future recall.

Denil reports a development of SKETCHPAD and SKETCHPAD III implemented on the IBM 7044 machine. The software is a *computer-aided design* (CAD) language which is controlled by keywords displayed in menus rather than on hardware function switches. Command keywords are displayed on the screen with vocabulary lists appropriate to the current situation, error messages, control options and the displayed object. The light pen is used to operate the vocabulary as light buttons. Points located by the light pen are acknowledged by drawing a square around them. Lines located are brightened. The general syntax is

<verb> <object>, <connector> <name>, <connector> <name> §

where § is a terminator. Typical verbs are

FORM CHANGE COPY DELETE VIEW

Objects are

LINE POINT SURF VIEW

Connectors are

FROM TO THRU PERPTO INTOF

the last two connectors being abbreviations for "perpendicular to" and "intersection of".

3.3.2

In the graphical representation of 3D objects, particularly bottles, pottery and other objects with spherical or cylindrical symmetry, it is often necessary to provide some illusion of depth by highlighting certain areas and shading others to varying degrees. Much of this work is based on the technique developed by S.A.Coons (1964,1967); the surfaces of the object are split up into defined

areas known as *Coons patches*. These patches have advantages in the design of space figures because they are well-suited for display and can conveniently be matched or made continuous with adjoining surfaces of any type. More precisely, their advantage arises from the fact that they are parametric and bounded. The use of parameters allows the easy definition of infinite real slopes as the ratio of two finite and computable numbers. This is important since areas of rapidly-changing slope often cause trouble in a mathematical definition. Parameters also allow the definition of three independent coordinate axes. Boundaries are defined with the aid of blending functions so that continuity with other surfaces across the common boundary may be defined up to any degree. Coons' 1964 paper does not go further than the use of cubics for the boundaries and blending functions. A cubic is the lowest order polynomial which can describe a space curve and contain a point of inflexion, and it would therefore appear to be suitable for design applications where continuity up to the first degree is all that is required. Both papers form part of Project MAC (Massachusetts Institute of Technology).

Wylie et al. (1967) describe an algorithm for the generation of half-tone perspective drawings of objects described by assemblies of triangular bounded planes. The results are toned pictures generated in a time which is competitive with the time necessary for plain line drawings. However, the source of illumination is assumed to be at the viewpoint, and since a point light source cannot "see" the shadows it casts, no shadows are rendered. This method is not very suitable for the representation of archaeological objects.

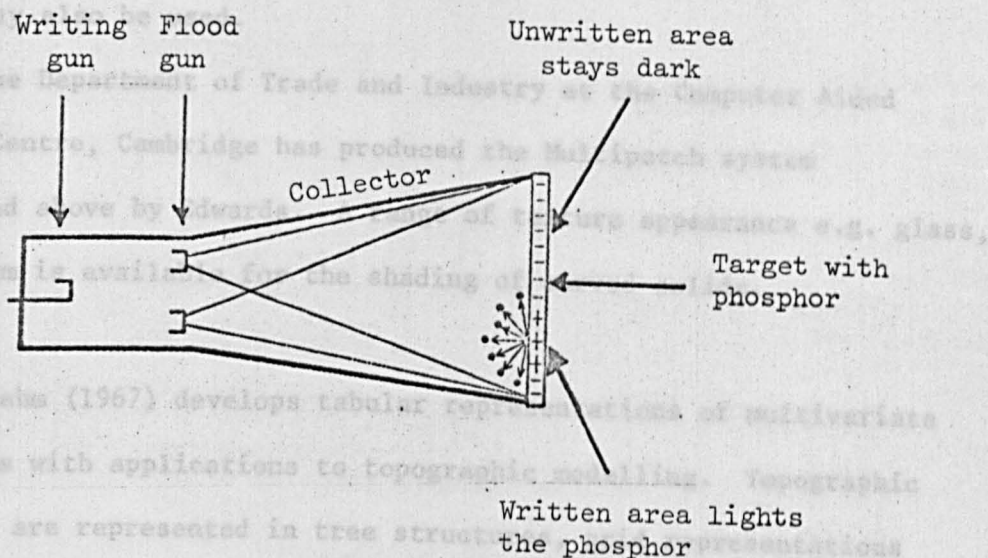
Appel (1968) gives some techniques for the machine rendering

of solids. Perspective drawings are rendered with hidden lines removed and toned shading, allowing shadows of deeper tone.

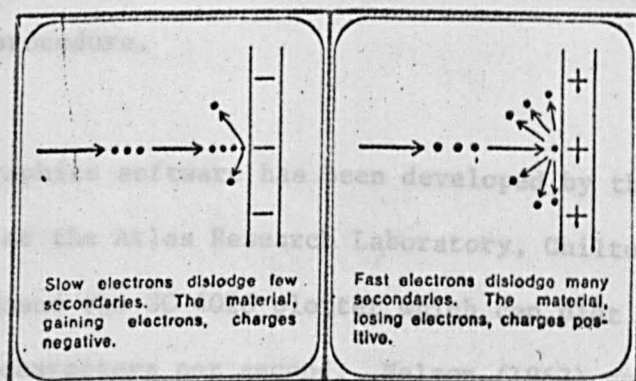
Armit (1969) describes a multipatch design system for Coons' patches, a powerful and very flexible tool using an interactive display terminal. The system appears to be based on the concept that the designer has some knowledge of the effect of altering surface coefficients, and also that he has some fore-knowledge of the object he is designing. An archaeological system must be far more automatic than this.

MacCallum (1970) describes the use of Coons' patches to design surfaces for display on a display terminal. Curves may be recognised by light pen and the indicated part of the curve "*dragged*" i.e. extended like a rubber band attached to the tip of the light pen until in the desired position. Modification of the curves is carried out in 2D and related to the 3D representation of rotated objects. The dragged segment should retain its end points and end slopes to ensure continuity of the first derivative. The usual choice of boundaries for Coons patches are cubic curves, but these are unsatisfactory for dragging and to give the extra degree of freedom quartic curves are necessary. To give more freedom in the dragging of local areas, the splitting of patches is employed. There is continuity across the common boundary of the split. Patches may also be joined; the principal patch does not change and the subsidiary patch is modified to give first-order continuity at the join.

Edwards (1971) describes the Cambridge Multipatch program using Coons' patches. The system may be used by remote terminal, using either refresh or direct-view bistable storage tube (DVBST, see Figure 3.2). A hard copy unit may be attached to the DVBST which scans the storage area of the tube when required, producing an



The flood gun electrons hit the unwritten areas too slowly to light the phosphor and the target charges negative. The positively-charged written area attracts electrons at high speed, keeping the phosphor lit and dislodging enough secondaries to hold the area positive.



The direct-view bistable storage tube

Figure 3.2

electrostatic copy of the diagram on paper. A hidden line algorithm may also be used.

The Department of Trade and Industry at the Computer Aided Design Centre, Cambridge has produced the Multipatch system described above by Edwards. A range of texture appearance e.g. glass, aluminium is available for the shading of curved solids.

3.3.3

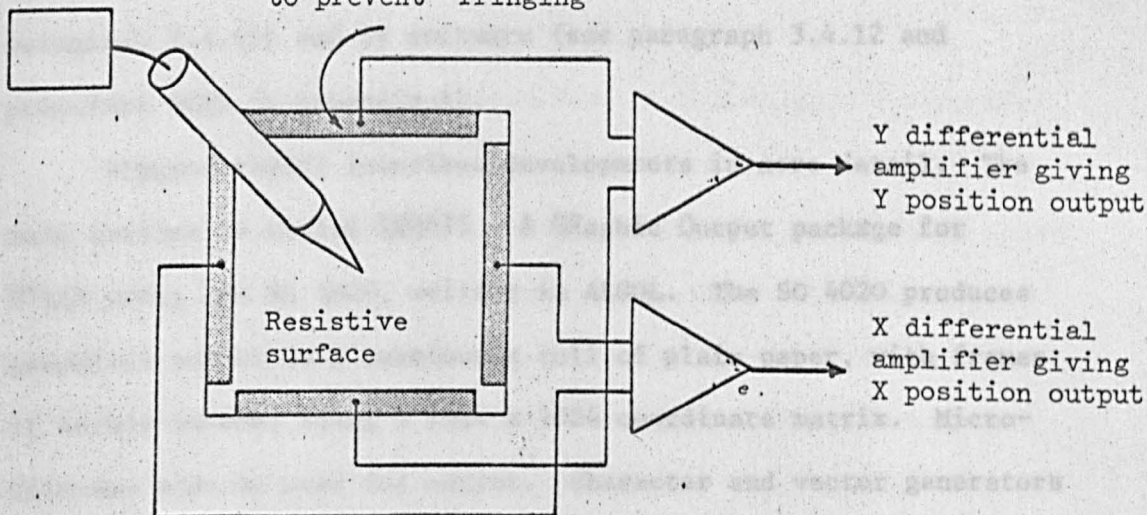
Boehm (1967) develops tabular representations of multivariate functions with applications to topographic modelling. Topographic surfaces are represented in tree structures, grid representations and contour representations. The *Rand tablet*, one of several types of tablet (using resistive, conductive and pulse-coded surfaces, see Figure 3.3) is used to define the coordinates of new points, replacing the light pen and its wasteful tracking routines. Other devices which similarly define new points are the *tracker ball* and the *joystick*, but these are not described by Boehm. The relationship between computer storage and speed in a graphics environment is explored. Boehm makes the point that graphics problems quickly become storage problems, and information retrieval problems. Boehm's variable grid method of representing multivariate functions seems to give the best results, but the encoding operation is a critical part of the procedure.

3.3.4

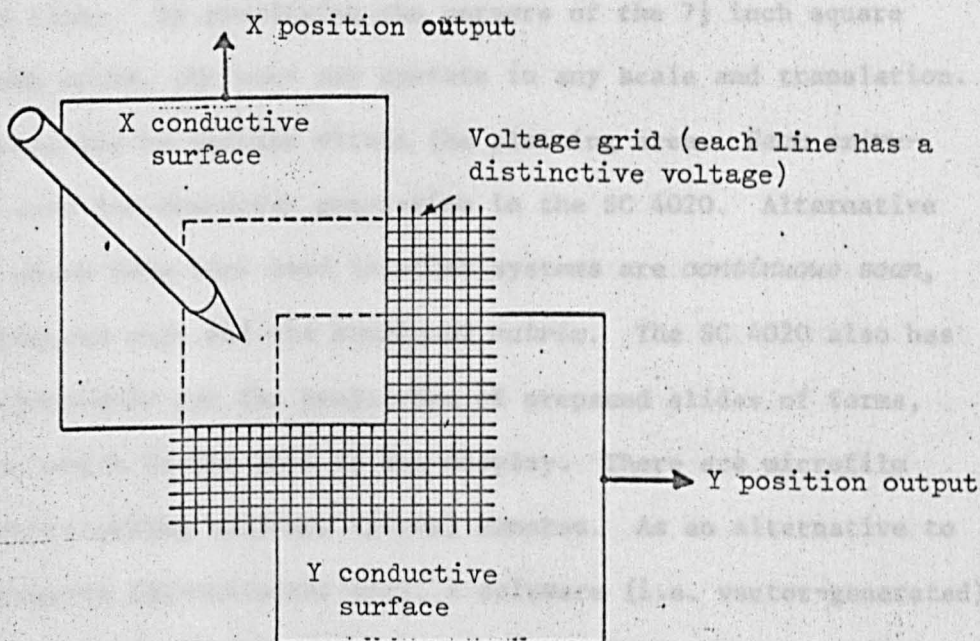
Graphics software has been developed by the Science Research Council at the Atlas Research Laboratory, Chilton. The system is built around the SC 4020 plotter which can plot up to 17,000 alphanumeric characters per second. Nelson (1967) reports FORTRAN software called HARTRAN. Examples are given of character sets generated by both hardware and software. The graphics routines developed for this thesis similarly use character sets generated by hardware (see

Voltage source

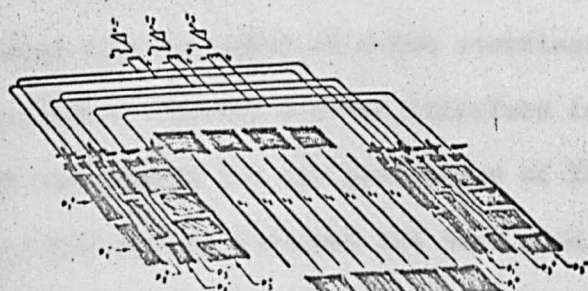
Linear conductive contact to prevent "fringing"



X-Y tablet using resistive surface techniques.



X-Y tablet using conductive surfaces voltage grid sandwich technique.



y_1	000
y_2	001
y_3	011
y_4	010
y_5	110
y_6	111
y_7	101
y_8	100

Coded pulses on lines
(1 positive, 0 negative)

A pen-like stylus picks up a time sequence of 6 pulses (positive or negative) indicating the position of the stylus

A pulsed tablet

Double-sided Printed-circuit Layout for 8 x 8 System.

Figure 3.3

paragraph 3.4.15) and by software (see paragraph 3.4.12 and procedure OCRB in Appendix A).

Hopgood (1969) describes developments in more detail. The main routine is called GROATS - A GRaphic Output package for ATLAS using the SC 4020, written in ALGOL. The SC 4020 produces graphical output on a continuous roll of plain paper, with frames $7\frac{1}{2}$ inches square, using a 1024 x 1024 coordinate matrix. Microfilm may also be used for output. Character and vector generators are available for the hardware generation of alphanumeric characters and lines. By specifying the corners of the $7\frac{1}{2}$ inch square in his own units, the user may operate in any scale and translation. Sub-regions may be defined within the plotting area. *Beam extrusion* is used for character generation in the SC 4020. Alternative methods which have been used in other systems are *continuous scan*, *non-continuous scan* and the *starburst matrix*. The SC 4020 also has a *forms projector* for the projection of prepared slides of forms, maps etc. which become part of the display. There are microfilm and photo-recording hardcopy (paper) cameras. As an alternative to beam extrusion (64-character set), a software (i.e. vector-generated) character set of 256 characters may be specified, which includes lower case letters, Greek letters, mathematical and plotting symbols.

Hopgood (1971) reports further developments, including a package for the input of d-Mac coordinates for the numerical specification of curves; a d-Mac interface to the PDP-15; and the programs CAMP and CAMPER for the production of 2D and 3D films respectively. The d-Mac pencil follower has been used frequently for the specification of curves in the programs developed for this thesis (see paragraphs 3.4.2, 3.4.3, 3.4.4, 3.4.9, 3.4.10, 3.4.11, 3.4.13, 3.4.15 and 3.4.18).

3.3.5

A number of graphics languages have been developed during the last five years. These are intended for the control of machine tools, graph plotters, automatic drafting machines, and display tubes.

Opitz and Simon (1967) formulate an extended version of the International APT (Automatic Programmed Tool) language, which originated in the U.S.A. and has been developed in West Germany and other European countries. The APT language has four classes of statement:

- (a) Organisational statements
- (b) Geometry statements
- (c) Motion statements
- (d) Work statements

Organisational statements reserve storage, define the part number, machine tool, and post processor (software which produces an individual control tape for the machine tool from the intermediate code common to all machines) and transform axes by rotation and translation. Motion statements move the cutting head of the machine tool around the workpiece. Work statements define the cutting operations. The statements chiefly of interest to this study are the geometry statements. A typical sequence is given below.

```
ZSURF/0.0
P1=POINT/24.0,16.0,0.0
L1=LINE/P1,(P2=POINT/10.0,12.0,0.0)
L2=LINE/P2,ATANGL,60
C1=CIRCLE/CENTER,P2,RADIUS,10
L3=LINE/P1,LEFT,TANTO,C1
L4=LINE/P1,RIGHT,TANTO,C1
```

This sequence defines the *z surface* of the workpiece, i.e. the default value of all future *z* coordinates. A point P1 is defined with

coordinates $x = 24.0$, $y = 16.0$, $z = 0.0$. A infinite straight line is defined which passes through P1 and another point P2 as yet undefined. Note that a *nested definition* is used for P2, which has coordinates (10.0, 12.0, 0.0). A line L2 is defined which passes through point P2 making an angle of 60° with the positive x axis. A circle C1 is defined with centre P2 and radius 10. Two more lines L3 and L4 are defined which pass through P1 and are tangential to C1, touching respectively the left and right sides of the circle as viewed from P1. This type of geometry is of limited use in archaeology, but a series of similar features have been provided in the graphics language developed for the plotter in this theses (see paragraph 3.4.6) and these are of use in drawing site plans, etc.

The Gerber Scientific Instrument Company (1969) have prepared the Gerber Graphics Generator (3G), which is high-level software for the operation of the Gerber automatic drafting machines, implemented on the IBM 360. The language describes precision artwork, and the program output is on cards or magnetic tape for off-line plotting on the automatic drafting machine; this consists of a photo-exposure head which produces a beam of light of variable aperture, intensity and duration. The plotting area consists of a photographic plate and the whole process is carried on in safelight conditions. The control tape moves the carriage and flashes the head to produce lines, pads and symbols as required.

Notley (1970) describes a graphical picture-drawing language designed specifically for use with the DVBST. He is of the opinion that there can only be a small number of applications in which the high cost of a refresh display is justified. The display of moving objects is only required in a few applications. Moreover, the DVBST does not require refresh, freeing the system for other tasks. The

software is implemented in NEAT for the ICL 4120 machine. The program uses the concept of a *drawing vector*, which has position, scale and orientation. *Move*, *draw* and *rotate* commands respectively move the drawing vector without modifying the display, draw a line and rotate the axes. Text is always written horizontally (this is probably a hardware limitation) as a "luggage tag" located at the correct place; Notley states that the number of applications where one would not wish the text to be horizontal and thus easily legible are small, but this is probably making a virtue out of necessity. Labels may be used in the command sequence, commands may be nested and there may be multiple repeats of a sequence of commands. A *stack* (first in - last out storage) is used to contain the drawing vectors which are nested. *Nest in* commands place the current drawing vector onto the stack, while *nest out* commands take a value off the stack and set the drawing vector to this value. A *base* command resets the drawing vector to the value at the top of the stack without nesting out. A *scale* command alters the scale of the subsequent commands to a percentage of the current value of the scale. There is storage for *global* and *local variables*, plus the current values for the coordinates, rotations and scale of the drawing vector. There are the usual arithmetic operators, names may be allocated to routines, *conditional expressions* may be used, and *recursion* is also possible. The formal syntax is given in Backus-Naur form. Matrix transformations are used for rotation of axes and perspective. Hard copy of any display may be obtained via the digital incremental plotter. Notley makes several suggestions about the way his program may be improved: addition of common functions to the syntax, text variables, drawing lines from the current position to an absolute position, and viewing the surface of curved objects.

Taylor (1971) gives a popular account of displays, light pens, tracker balls, tablets, *touchwires* (which are wires located in a plastic mask on the face of the tube; characters may be displayed behind them, and a signal is sent to the computer when a wire is touched by the finger, thus touchwires have the same purpose as hardware function switches), microfilm devices, electrostatic printers, display files and DVBSTs. He describes current display software, e.g. IBM GSP; ICL FRED, EDGAR and DISMAN. There is a survey of proprietary displays in use in the UK. This thesis uses the display language EDGAR, which is an extension of ALGOL 60 implemented on the ICL 4100 range of machines.

3.3.6

Digital incremental plotters have a *drum* which carries the roll of paper, and which may be stepped single increments (e.g. 0.01 cm, 0.005 inch) by a *stepping motor*. A *pen carriage* moves a pen above the paper, and the pen may be moved (usually at right angles to the drum motion) in similar increments. The pen may be raised, or lowered to touch the paper when it will draw a line. Thus simple plotters are restricted in motion to eight vectors, composed of single increments, either positive or negative, of the drum or pen carriage, or both simultaneously. More sophisticated plotters have 16 different directions, two different increment sizes and variable speed. In simple plotters the directions are calculated by a *software* algorithm, and a "straight" line is the zig-zag which most closely fits the required direction. In more sophisticated plotters *hardware* may direct the plotting, and vector directions, increment sizes and plotting speeds are chosen to fit the curvature of the line being drawn - a nearly straight line will have large increments and fast speed, while a tight curve will have small increments and slow speed. Most installations

have simple plotters, since the large sophisticated plotters are extremely expensive, and thus much computer time is spent in directing the motion of the pen for even the simplest straight line. For more complex curves the time overhead may be prohibitive, so much research has been instituted to find efficient algorithms for the common curves such as conic sections. Pitteway (1967) describes such an algorithm for the plotting of ellipses and hyperbolae.

3.3.7

When the illusion of 3D is to be given by means of a 2D display, perspective calculations are performed. This, however, is not enough because a "wireframe" representation of an object does not look real. To make the object look solid it is necessary to remove lines representing edges of the object which should be hidden from the observer's viewpoint (the "hidden line" problem) and to use graded shading. Several papers have been devoted to the "hidden line" problem.

Comba (1968) describes a procedure for detecting intersections of 3D objects. This has applications in the detection of hidden lines.

Treece (1970) gives a technique for the removal of hidden lines. Three mutually-orthogonal planes P, Q and R are defined, each containing the origin of the coordinate system. The intersection of planes Q and R is fixed arbitrarily. To find the coordinates of any point in an object in the new axes defined by the intersections of the planes, perpendiculars are dropped to the three planes. Q is taken as the plane of the screen and contains the projected 2D image of the object. Any intersections of lines in this image are checked back to the equivalent points in 3D, and the line which is the more remote from the observer is subdivided into two segments. In a realistic

view the more remote segments which would be obscured by nearer surfaces are deleted. This is a simple approach, intended as a first step towards justifying a more sophisticated program.

Jones (1971) reviews past methods of solving the hidden line problem and classifies them as:

(a) more or less complete comparison of each polygonal face with each line to determine overlap;

(b) scan of the objects with a cutter plane passing through the viewpoint, on which intersections are determined, and a raster image built up;

(c) repeated division of the picture into quadrants until a portion is found which may be processed simply. Jones proposes a new method which describes 3D objects as a series of interconnected cells. The cells are bounded by faces which are either the faces of the objects or artificial transparent faces. All cells are convex polyhedra and may have opaque or transparent faces. A stack is used to store the successive masks representing areas of visibility from a particular viewpoint, e.g. a view through a door may disclose a room with a serving hatch giving onto a second room. Successive masks representing the visible areas of the first and second rooms are the door frame and the frame of the hatch respectively. Data words are used to describe cell-elements and opening-elements. Connected graphs are used to describe lines of sight.

Ricci (1971) gives an algorithm for the removal of hidden lines. Objects are considered as sets of lines, comprising polyhedra, polygon-bounded planar surfaces and straight line segments. Every object is compared with all segments in the scene. Any hidden parts are removed by removing a complete segment from the segment list, or substituting one or more segments corresponding to the part or parts of the lines left visible. The algorithm is implemented in FORTRAN.

Hidden line removal is not treated in this thesis except insofar as an unwanted line may be deleted from the display by light pen, and a small segment replaced if necessary. Another solution, which is used in ARCJW025 (see paragraph 3.4.16) for the display of a sectioned site excavation in pseudo-3D perspective, is to input only two adjacent side views and the top view of the rectangular block of earth, and rotate the view so that the remaining two sides and bottom would be obscured in any case. This latter method is probably only applicable to cuboidal structures, such as the archaeological example given above, but it has not been seen elsewhere. The efficacy of the technique may be judged from Figure 3.21.

3.3.8

All display applications need alphanumeric characters for the display of legends or text. The required characters may be generated either by hardware or software, as has been described above in paragraph 3.3.4. The hardware-generated characters are usually limited in number and format, particularly if *beam extrusion* is used, since the characters are then generated by an internal mask which forms the beam of electrons into the desired shapes. The *continuous scan* and *non-continuous* scan methods are less restrictive, each character being generated as a series of matrix dots or a series of incremental segments. Because of speed considerations, however, hardware-generated characters tend to be more or less stylised - this may be judged from the legends on the display-generated Figures in this chapter (the ICL 4100 series display uses non-continuous scan in an incremental mode for the hardware-generation of characters). Software generation of characters is much slower, but the programmer then has a free hand in the design of a pleasing font or fonts, and such characters as Greek letters, mathematical and plotting symbols can easily be included.

Such a system is described by Heap and Laws (1968). The construction of the characters available in the National Physical Laboratory graphical output system (English Electric KDF9) is given. Numerals, Roman and Greek alphabets (upper and lower case), punctuation signs, mathematical symbols and point plotting symbols are available; where possible, these have been designed to follow closely the ISO OCRB font for character recognition use. The font is pleasing to the eye. The characters are specified in octal on a 4 x 7 grid, codes 00-47 specifying coordinates; code 5Y causes the value Y to be added to all y coordinates until further notice, thus raising the character, and code 6Y similarly subtracts Y from all y coordinates, lowering the character; code 7Y causes the plotter pen to be raised, and moved to the next coordinate. Characters may be italicised by progressive displacement of the x position as the y displacement increases, using the equation:

$$x_i = x + 0.2y$$

which gives a slope of about 11°. A similar character set, but based on the ICL 4130 internal code, and requiring the design of new characters has been written in ALGOL 60 for this thesis (see procedure OCRB, Appendix A and paragraph 3.4.12). Variable italic slope, and orientation of the whole text are additional features which have been incorporated (see Figure A2, Appendix A).

3.3.9

Luzadder (1968) in a general textbook on graphic techniques, mainly for the draughtsman, describes the design of graphic output for the most effective man-machine communication. A nomenclature is given to describe features of perspective drawings, e.g. *picture plane*, *horizon line*, *axis of vision*, *ground line*, *horizon plane*, *ground plane* and *vanishing point*. The connection between computer-aided design and

automatic drafting is explored. There is a good survey of graphics hardware. Microfilm slides may be stored on aperture cards, and by placing such a card in a scanning device, the diagram may be converted to digital data. Interactive graphics, the microfilm printer/plotter, automatic drafting machines and graphics languages are described. The general forms of diagrams and charts are discussed, e.g. piecharts, histograms, three-pole plots and geographical "stack" charts (where stacks of items, or the heights of rods indicate the values of some variable at various geographical locations). All these have applications in archaeology.

3.3.10

Patrick, Anderson and Bechtel (1968) describe an algorithm for mapping multidimensional space to one dimension for computer output display. The problem considered is the mapping of a function's multidimensional bounded domain onto a bounded interval of a real line. This transformation is to have the property that "neighbouring" points in the bounded interval are necessarily mapped from "neighbouring" points in the bounded domain. The multidimensional bounded domain is partitioned into a finite number of elementary regions while the bounded interval is partitioned into the same finite number of elementary intervals. A one-to-one correspondence is defined between the elementary regions and intervals such that neighbouring elementary intervals have corresponding multidimensional elementary regions which are also neighbouring. Two types of mapping are described, dovetail mapping and column mapping. The intervals between points on the line are proportional to the volumes of the regions. This technique is applicable to the method of multidimensional scaling described in the statistics chapter (section 1.6).

3.3.11

Shepard (1968) describes a two-dimensional interpolation function

for computer mapping of irregularly-spaced data. There is a need for interpolation of irregularly-spaced data to produce a continuous surface. In archaeology, these irregularly-spaced points may be excavation sites or survey locations. In order to display this data in some type of contour map or perspective view, to compare the set of observations with other information for the same region based on different sampling points, or to analyse the data for extremes of value, gradient, etc., it is useful, if not essential to define a continuous function fitting the given data exactly. Interpolated values over a fine grid may then be defined. The paper describes such a function, and gives examples of symbol maps produced on the line printer.

3.3.12

One method which has been used on computer displays to produce an illusion of depth is to display a stereo pair and view them through a special optical arrangement, using the stereoscopic property of the eyes to give depth perception.

Schoenaker (1968) describes such a system, producing stereo pairs in two different colours on the plotter using the COBRA machine, which was specially designed for use as a universal controller of machine tools, but which can also control plotters. The results are viewed through the time honoured spectacles with different-coloured lenses.

Ortony (1971) produces a stereo pair on a display screen, and uses an optical method of combining the images which allows several people to view simultaneously. The two images use horizontal and vertical polarisation, which are transmitted and reflected by a transmitting/reflecting surface. Polarised spectacles are used to view the 3D reconstruction. The reflected image must be horizontally

polarised because of the polarisation properties of a reflecting surface. A mask is arranged to obscure the upper image on the screen, which would otherwise be directly visible by one eye. One image, the upper one on the screen, is inverted because of the need for reflection of this image. Since it is inconvenient to remove the viewing assembly for a series of photographs, it is better to store the display files for a series of stereo pairs, and reproduce these later as conventionally-oriented pairs for photography.

3.3.13

Wilcock (1968) compares computer graphic techniques for the automatic reduction of archaeomagnetic field observations, with particular reference to the computer-archaeologist interface and symbol diagrams produced by line printer. Symbol diagrams in general are not very successful - this partly springs from the different horizontal ($\frac{1}{10}$ inch) and vertical ($\frac{1}{6}$ inch) spacings on most printers, making it difficult to obtain compact diagrams without areal distortion, and partly from difficulties in perception. Dot-density maps produced on the plotter are much better, and these are described in the Survey Reduction chapter, paragraph

3.3.14

Biddle and Kjølbye-Biddle (1969) give the first detailed suggestions for the use of the metric system in British archaeology (apart from a brief note by R.J.C. Atkinson, 1965). The metric system in 3D is essential for the computer recording of archaeological finds on site.

3.3.15

Hall, Ball, Wolf and Eusebio (1969) describe an interactive graphics pattern-recognition system, PROMENADE, implemented in FORTRAN on the CDC 3100 machine. A "mouse" i.e. a type of pencil follower used to input new points (replacing the light pen) and keyboard are used for man-machine interaction. "Metroglyphs", a way of displaying

the values of more than one variable at a single point on the screen are used. In this method the values of two of the variables are represented by x and y coordinates, while the length of a line starting at the point may represent the value of the third variable, and the slope of the line the value of the fourth variable. The system performs clustering of data points and constructs link-plots for pattern recognition.

3.3.16

Rosing (1969) describes the construction of dot or symbol maps using overprinting. Disadvantages of this technique are the different spacings on most printers in the horizontal and vertical directions, the lack of an overprinting facility on many printers, and the production of a less aesthetically pleasing diagram than a dot-density map. In Rosing's application symbols form the outline and body of the country portrayed in the map, and the density of the symbols may represent, for example, population density. Software called LINMAP has been developed by the Urban Planning Directorate III of the Ministry of Housing and Local Government. Alternatively the symbols plotted may be numerals, and these may form the basis for a hand-isopleth map, e.g. a contour map. *Choropleth maps* represent areas of known boundaries as having equal density of a variable across their entire surface. Trend surface maps (see section 1.8) may be plotted as dot maps on the line printer, with representation of the residuals. The use of the plotter is briefly mentioned.

3.3.17

What appears to be the first archaeological publication describing the use of the display unit is a paper by Virginia Burton (1970), Assistant Curator of Egyptian Art, Metropolitan Museum of Art. She describes freehand drawing of figures on the cathode ray tube using

the light pen. Using this method Egyptian pottery is classified according to shape and decoration, although few details are given of the actual methods used. The work of Flinders Petrie in Egyptology is reviewed; it is mentioned that British workers have adhered to his standards, but workers in other countries have used other systems of classification, or sometimes no system at all. The use of a tablet to input line drawings of pots, by laying a drawing or photograph on the tablet, or projecting a slide onto it, and following its outline with the stylus (see Figure 3.3), is described. This is a more direct, but similar method to the d-Mac pencil follower method employed in this thesis for pottery classification. Point-to-point measurements are communicated directly by identifying the points on the line drawing; the action is acknowledged by a straight line drawn on the screen between the two points. The d-Mac method of this thesis provides coordinates of a large number of points which are simultaneously available to the computer for the calculation of various measurements. Light buttons are used to scale the diagrams, to correct foreshortening of the photographs used for input, to change the size of part of the diagram, to compare shapes, and to present a rotating 3D display calculated from the plane views. The limited number of hardware function switches may be used for several different functions each by using one key to step between several different tables of functions, in rotation. This is a very useful technique, but in this thesis it was decided to retain commonly-used functions for the 8 hardware function switches available on the ICL 4100 display unit and to extend the range of functions by the use of light buttons and teletype input. The use of the display unit format for direct publication is described; that this is a great advantage to the archaeologist is a

main claim of this thesis. The Metropolitan Museum of Art system uses a *Geo-space plotter* which produces diagrams of size 40 inches by 60 inches in 32 shades of grey, taking 75 seconds per print. Not many installations could afford such an expensive facility; the simple digital incremental plotter is adequate for the production of hard copy from the display file. The use of *overlays* for the hardware function switches is described. These are plastic masks with legends identifying the function of each key, or alternatively transparencies with symbols which are back-illuminated by the computer when a response by function key is required by the software.

3.3.18

The University of Arizona CDC 6400 machine has been used for archaeological applications (1971). A list of programs frequently used by archaeologists is published, including the following:

(a) PLOT3D. This program draws 3D histograms and surfaces (including mathematical functions) viewed from any angle on the Calcomp digital incremental plotter.

(b) SYMAP. This program performs synagraphic computer mapping, producing maps which graphically depict spatially disposed quantitative and qualitative information. It is suited to a broad range of applications and is provided with numerous options to meet varying requirements, including a contouring capability.

3.3.19

d-Mac Limited describe the *d-Mac Pencil Follower Type P.F. 10,000 Mark 1B*. This is the type of pencil follower which has been used for this thesis, and is illustrated in Figure 3.4. Unfortunately the *model 1B* is now becoming obsolete, and d-Mac are promoting the *model 2* which has adjustable origin, 5-figure accuracy and a *menu* for complete interaction via a *CAMAC* interface with a small computer

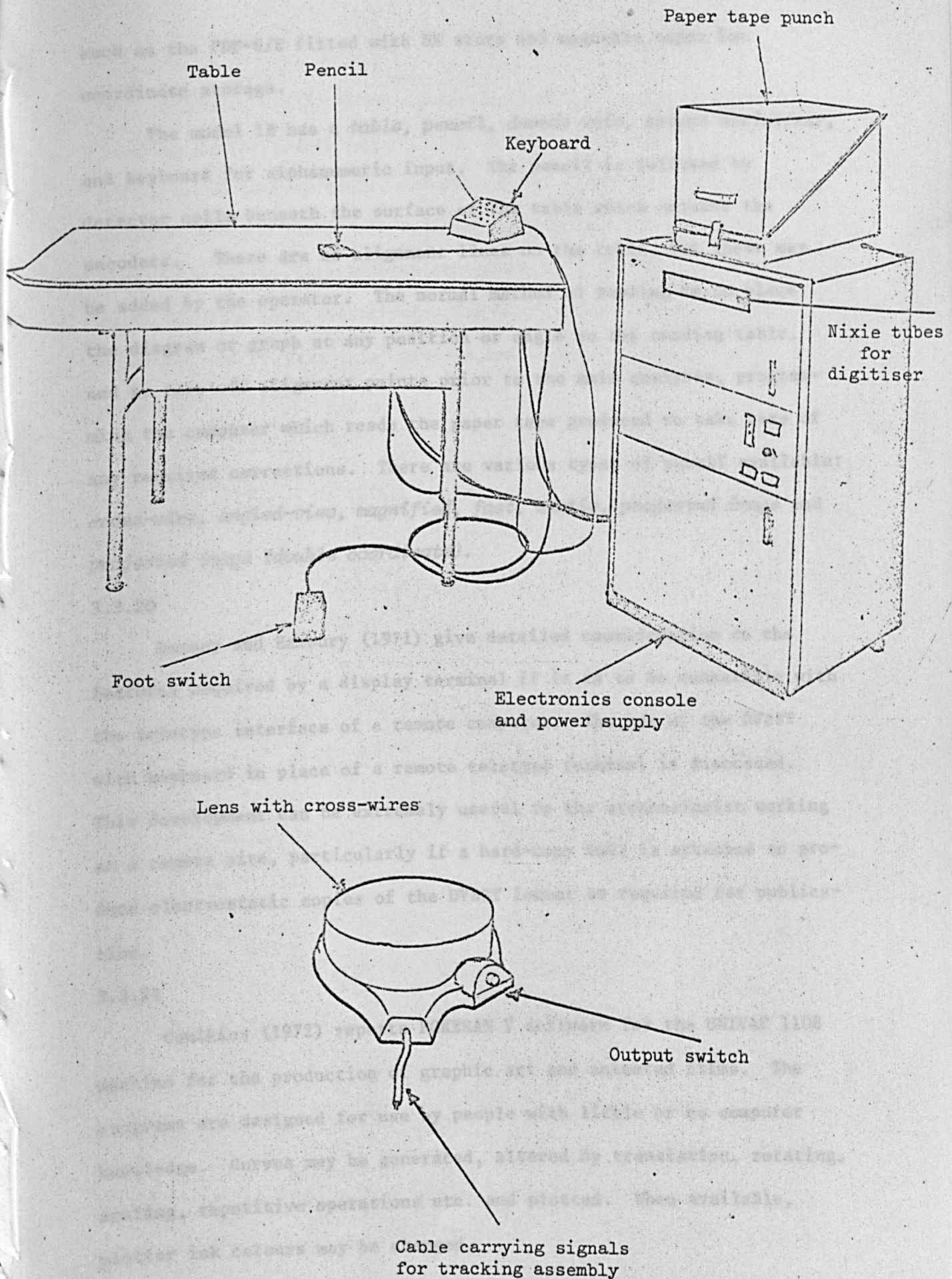


Figure 3.4

D-mac Pencil Follower 1B

such as the PDP-8/E fitted with 8K store and magnetic tapes for coordinate storage.

The model 1B has a *table, pencil, decode unit, output serialiser,* and keyboard for alphanumeric input. The pencil is followed by detector coils beneath the surface of the table which actuate the encoders. There are no alignment lines on the table, but these may be added by the operator. The normal method of reading is to place the diagram or graph at any position or angle on the reading table, and to take off alignment points prior to the main analysis, programming the computer which reads the paper tape produced to take care of any required corrections. There are various types of pencil available: *cross-wire, angled-view, magnified, fast, stable, projected image and projected image (double coordinate).*

3.3.20

Barney and Hambury (1971) give detailed consideration to the features required by a display terminal if it is to be compatible with the teletype interface of a remote computer. The use of the DVBST with keyboard in place of a remote teletype terminal is discussed. This development can be extremely useful to the archaeologist working at a remote site, particularly if a hard-copy unit is attached to produce electrostatic copies of the DVBST format as required for publication.

3.3.21

Caulkins (1972) reports FORTRAN V software for the UNIVAC 1108 machine for the production of graphic art and animated films. The programs are designed for use by people with little or no computer knowledge. Curves may be generated, altered by translation, rotating, scaling, repetitive operations etc. and plotted. When available, plotter ink colours may be changed.

3.3.22

Pitteway (1972) gives a general summary of computer graphics covering the following topics: line printer, teletype, microfilm, plotters, NC machine tools, cathode ray tube, DVBST, stereo viewing devices, tablets and pencil follower. Techniques for providing grey tones on the line printer are given (. / X M for increasing depth of tone) with overprinting for black tones. Electrostatic printers, the "hidden line" problem and animated films are also discussed.

3.4 *Graphics programs developed for this thesis*

3.4.1

*ARCJW001 Archaeological plotting generator
(ALGOL 60)*

This program performs a dot-density plot of an archaeological site. The survey process is described in the Survey Reduction chapter (chapter 5). The program reads data from cards, consisting of the site name, dimensions of the site, a mode parameter (random or systematic plot, usually the former) and the electrode or reading spacing. Each set of survey squares is identified by a reference number, and within each set each square is similarly identified. A survey square is made up of a series of cards, each card carrying the readings for a line of the survey and identified by a line number. The line numbers are checked automatically to detect any omissions in the readings. The readings are manipulated to enhance anomalies - this is a simple form of *filtering* (see chapter 5). For Banjo meter readings it has been found useful to square all readings, then divide by 20. The nearest integer is then used as the dot count for the particular reading. A random number generator (procedure RANNO, see Appendix A) is used to position dots randomly within a square of side equal to the reading spacing, and centred on the reading station.

On request from Leslie Alcock, director of the South Cadbury excavations, plots were initially randomised within a circle of radius $a\sqrt{2}/2$ (where a is the reading spacing) centred on the reading station, in order to represent the approximate range of the Banjo meter, but this gave a misleading plot (see references Alcock, L., 1967-1972 and Wilcock, J.D., 1970 *Prospezioni Archeologiche* 5, 55-58, in the Survey Reduction Bibliography). A typical dot-density plot produced by the program is illustrated in Figure 3.5. The program uses procedures PLOT (plots a point at the specified location as a small square of side 0.02 inch), RANNO (random number generator) and FILTER (squares readings, divides by 20 and finds the nearest integer as a dot count - designed for the Banjo meter). All these procedures are described in Appendix A.

3.4.2

ARCJW006 Transcriber for card data and paper tape profiles (ALGOL 60)

This program allows profiles of objects, e.g. clay pipes to be transcribed to the plotter as well as written to magnetic tape. An example of the output is given in Figure 2.3. The program is fully described in the Information Retrieval chapter, paragraph 2.6.2.

3.4.3

ARCJW007 Map transcriber for distribution maps (ALGOL 60)

This program reads d-Mac pencil follower coordinates specifying a map of the British Isles compiled of several sections i.e. the mainland, Ireland, Lough Neagh, and major offshore islands, many of which are measured relative to different grid origins. The different sections of the map are rotated and translated as necessary, then recorded on magnetic tape for subsequent plotting. This program was custom-built for the map of the British Isles related to the origins of the National Grid and the Irish Grid (which is not marked on most

readily-available large scale maps of Ireland). A more general map generator, ARCJWO22, see paragraph 3.4.14, has subsequently been developed.

3.4.4

ARCJWO22 MAPGRI Fast program (ALCOL 60)

This is a skeleton program to test the procedures MAPGRI and SHAPE which respectively draw a map outline and Ireland, and shade specific areas of the map. It is a custom-built procedure which plots a map from the magnetic tape coordinates generated by ARCJWO27 (see paragraph 3.4.13). Each call of the procedure may draw a new map or use the map drawn by a previous call of the procedure. If a new map is drawn, a title consisting of a variable number of strings is added, together with a grid north-point, bearing and scale in 24. Points may be plotted on a map at specified National Grid References (each consisting of two letters plus six figures) using various plotting options: solid, dashed, dotted, dot and doubly-enhanced dot. The map is plotted with respect to the British National Grid, as used for Ireland. The map specification is read from paper tape, while the map coordinates are read from magnetic tape. Alternatively, the grid references may be taken from an array. Typical maps using several different types of symbols are shown in Figures 2.7 and 3.4.

SHAPE all SOUTH CADBURY 1969
grid references, to be shaded in a variety of ways. The grid references specifying the map outline of the area to be shaded may be read from paper tape or from an array. The following shade elements which build up to form the shading pattern are available: +, diamond, X,

dot and cross-hatching
Figure 3.5
Dot-density plot

readily-available large scale maps of Ireland). A more general map generator, ARCJWO22, see paragraph 3.4.14, has subsequently been developed.

3.4.4

ARCJW009 MAPGBI Test program (ALGOL 60)

This is a skeleton program to test the procedures MAPGBI and SHADE which respectively draw a map of Great Britain and Ireland, and shade specified areas of the map. MAPGBI is a custom-built procedure which plots a map using the magnetic tape coordinates generated by ARCJW007 (see paragraph 3.4.3). Each call of the procedure may draw a new map or use an existing map drawn by a previous call of the procedure. If a new map is drawn, a title consisting of a variable number of strings is added, together with a grid north point, border and scale in Km. Points may be plotted on the map at specified National Grid References (each consisting of two letters plus six figures) using various plotting symbols: +, diamond, X, dot, enhanced dot and doubly-enhanced dot. The Irish Grid, which is tilted with respect to the British National Grid, is used for Ireland. The map specification is read from paper tape, while the map coordinates are read from magnetic tape. Alternatively, the grid references may be taken from an array. Typical maps using several different types of symbols are shown in Figures 2.7 and 3.6.

SHADE allows closed areas of the map, specified by chains of grid references, to be shaded in a variety of ways. The grid references specifying the map outline of the area to be shaded may be read from paper tape or from an array. The following shade elements which build up to form the shading pattern are available: +, diamond, X, dot and cross-hatching in the vertical, positive diagonal, horizontal and negative diagonal senses. An algorithm similar to a maze-searching

COMPARISON OF DISTRIBUTIONS OF IRON AGE BARROWS (.) AND BARROW GROUPS (+) WITH ROMAN BARROWS AND MAUSOLEA (X)

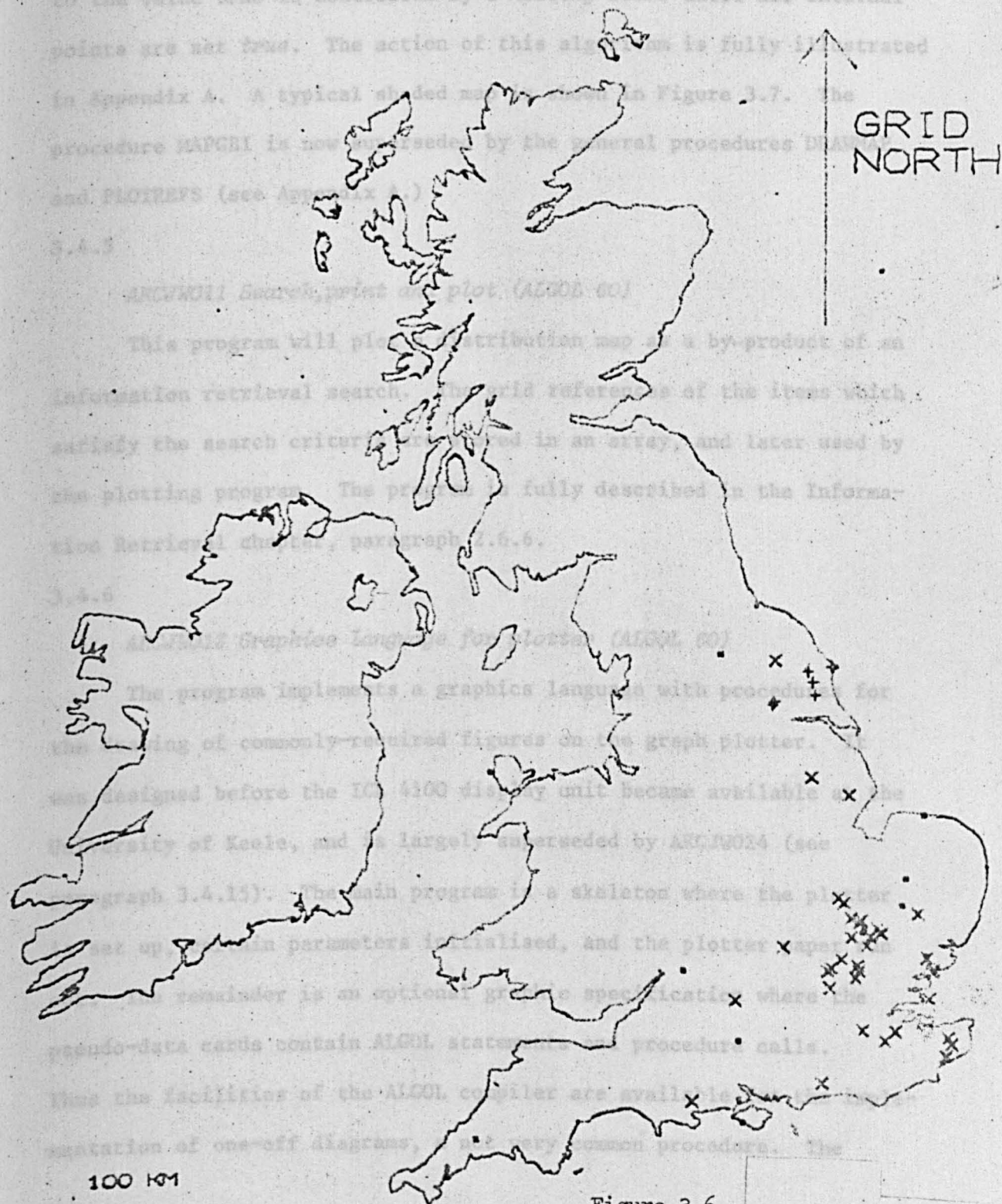


Figure 3.6

routine is defined to determine which of the available matrix points at 0.1 inch spacing shall be shaded. The matrix points are all set *false*, then the boundary of the area is defined by a connected series of *true* points. Starting at an interior point, also specified by paper tape or in an array, the setting of the boolean matrix points to the value *true* is controlled by a *nesting store* until all internal points are set *true*. The action of this algorithm is fully illustrated in Appendix A. A typical shaded map is shown in Figure 3.7. The procedure MAPGBI is now superseded by the general procedures DRAWMAP and PLOTREFS (see Appendix A.)

3.4.5

ARCJW011 Search, print and plot (ALGOL 60)

This program will plot a distribution map as a by-product of an information retrieval search. The grid references of the items which satisfy the search criteria are stored in an array, and later used by the plotting program. The program is fully described in the Information Retrieval chapter, paragraph 2.6.6.

3.4.6

ARCJW013 Graphics language for plotter (ALGOL 60)

The program implements a graphics language with procedures for the drawing of commonly-required figures on the graph plotter. It was designed before the ICL 4100 display unit became available at the University of Keele, and is largely superseded by ARCJW024 (see paragraph 3.4.15). The main program is a skeleton where the plotter is set up, certain parameters initialised, and the plotter paper run out. The remainder is an optional graphic specification where the pseudo-data cards contain ALGOL statements and procedure calls. Thus the facilities of the ALGOL compiler are available for the implementation of one-off diagrams, a not very common procedure. The

SHADED AREAS FOR THE COUNTIES OF YORKSHIRE, NORFOLK, DEVON AND CLARE



MACROS

Reads a G-Mac tape specifying a curve

LIMITS

Checks that the required point is within the limits of the diagram

POSITION

Positions the pen at the desired point, in pen-up position; uses LIMITS

ORIGIN

Redefines the origin of the diagram; uses POSITION and SETORIGIN

READFIG

Reads a curve and normal to the curve which specifies

DRAWFIG

Draws a curve and normal to the curve which specifies

PERM

A recursive procedure to produce output procedures which are called by the data code which the procedure body; for graphical output the procedure is to generate their position and fill coding. Figure A4. This procedure has been used to produce diagrams of strike-slip faults for possible chronological sequences

WRITE

Writes text as a string

INTNUM

Writes a decimal integer

REALNUM

Writes a real number, integer and fractional parts separated by a decimal point

SCALE

Alters the scale of the diagram

SPECIFY

Specifies a circle by explicit coordinates

ALIAS

Specifies a circle by radius and center

SPECIFY

Specifies a circle by explicit coordinates

SPECIFY

Specifies a circle by radius and center

SPECIFY

Specifies a circle by explicit coordinates

SPECIFY

Specifies a circle by radius and center

SPECIFY

Specifies a circle by explicit coordinates

TRANS

Translates the origin of the diagram

ZSURF

Alters the Z surface

ROTATE

Rotates the diagram

GRID
NORTH

100 KM

Figure 3.7

following ALGOL procedures may be called (all are fully described in Appendix A):

DMACIN	Reads a d-Mac tape specifying a curve
LIMITS	Checks that the required point is within the limits of the diagram
POSITION	Positions the pen at the desired point, in pen-up position; uses LIMITS
ORIGIN	Redefines the origin of the diagram; uses POSITION and SETORIGIN
READFIG	Reads a d-Mac tape and normalises the curve which it specifies
DRAWFIG	Draws a curve previously read in by READFIG
PERM	A recursive routine to permute output procedures which are called by pseudo-data cards within the procedure body; for graphical outputs the intention is to permute their positions. The full coding is given in Appendix A, Figure A3, and a typical output in Figure A4. This procedure has been used to permute diagrams of strike-a-lights for Dr F.Celoria into all possible chronological sequences
WRITE	Writes text specified as a string
INTNUM	Writes a decimal integer
REALNUM	Writes a real number, with integer and fractional parts separated by a decimal point
SCALE	Alters the scale of succeeding commands
ARCCO	Specifies an arc by explicit coordinates
ARCSYM	Specifies an arc by symbolic names
CIRCCO	Specifies a circle by explicit coordinates
CIRCSYM	Specifies a circle by symbolic names
CTPTSSR	Specifies a Circle Through Points and of Specified Radius
LINECO	Specifies a line by explicit coordinates
LINESYM	Specifies a line by symbolic names
TRANS	Translates the origin of the diagram
ZSURF	Alters the Z surface of the diagram
ROTATE	Rotates the axes of the diagram

SIZE	Defines the extent of the diagram
POINT	Specifies a point
DRAWA	Draws a previously-specified arc
DRAWC	Draws a previously-specified circle
DRAWL	Draws a previously-specified line
GRID	Draws an archaeological site grid

A typical coding sequence specifying a diagram is given in Figure 3.8 and a generated diagram showing some of the capabilities of the system in Figure 3.9.

3.4.7

ARCJW014 Histogram test program (ALGOL 60)

This is a skeleton program which tests the procedures HISTOGRAM and PIECHART, to generate frequently-used types of diagram in archaeology. The program initialises parameters, uses procedure SCRIPT to read a number of title strings (which may use both upper and lower case letters, see Appendix A and Figures 3.12 and 3.13), then a definition of the readings about to be input, e.g. WEIGHT IN GRAMS. All these strings are printed on the line printer. Next the procedure PREHIST is called, which effectively turns the procedure HISTOGRAM into a "lazy man's histogram" for which the number of items need not be counted, nor the range and interval specified. Readings are input until an out-of-range value terminates this phase. The readings are printed, and the maximum, minimum and range found. The interval for the histogram is then automatically defined as a function of the range, and the maximum and minimum values are adjusted to be at interval points containing the full range of the readings.

Procedure HISTOGRAM uses the parameters calculated by PREHIST to plot a histogram of the distribution of the readings into the selected intervals. The height of the columns is scaled automatically. The x axis may be suppressed if desired (see the use of HISTOGRAM in


```

"COMMENT"GRAPHICS SPECIFICATION;
POINT(P1,-8.0,-8.0,0,0);
GRID(P1,0,4.0,4,4);
SCALE(0:1);
POINT(P2,-80.0,-80.0,0,0);
GRID(P2,0,4.0,10,10);
SCALE(1:0);
LINECO(L1,-5.0,-1.0,0,2.0,6.0,0,0);
LINECO(L2,4.0,6.0,0,7.0,3.0,0,0);
LINECO(L3,7.0,1.0,0,0,-6.0,0,0);
LINECO(L4,-2.0,-6.0,0,-5.0,-3.0,0,0);
ARCCO(A1,-4.0,-2.0,0,1.4142,135,225,0);
ARCCO(A2,3.0,5.0,0,1.4142,45,135,0);
ARCCO(A3,6.0,2.0,0,1.4142,-45,45,0);
ARCCO(A4,-1.0,-5.0,0,1.4142,225,315,0);
DRAWA(A1);DRAWL(L1);
DRAWA(A2);DRAWL(L2);
DRAWA(A3);DRAWL(L3);
DRAWA(A4);DRAWL(L4);
WRITE(-8.0,-9.0,8,'SITE GRID AND ROMAN CAMP');
INT1:=123;REAL1:=456.7890;
INTNUM(-8.0,9.0,3,4,INT1);
REALNUM(0,9.0,3,4,6,REAL1);
CTPTSSR(C1,8.0,8.0,0,8.0,-4.0,0,7.2111,-1,0,0);
DRAWA(C1);
CTPTSSR(C2,-6.0,8.0,0,-6.0,4.0,0,2.0,1,0,0);
DRAWC(C2);
ORIGIN(12.0,4.0);
POSITION(0,0,PLOT);
PLOT:CENCHARACTER(1);
"COMMENT"END OF GRAPHIC SPECIFICATION;

```

Large grid

Small grid

Completion of specification
of Roman camp outline

Drawing commands for the
Roman camp outline

Title

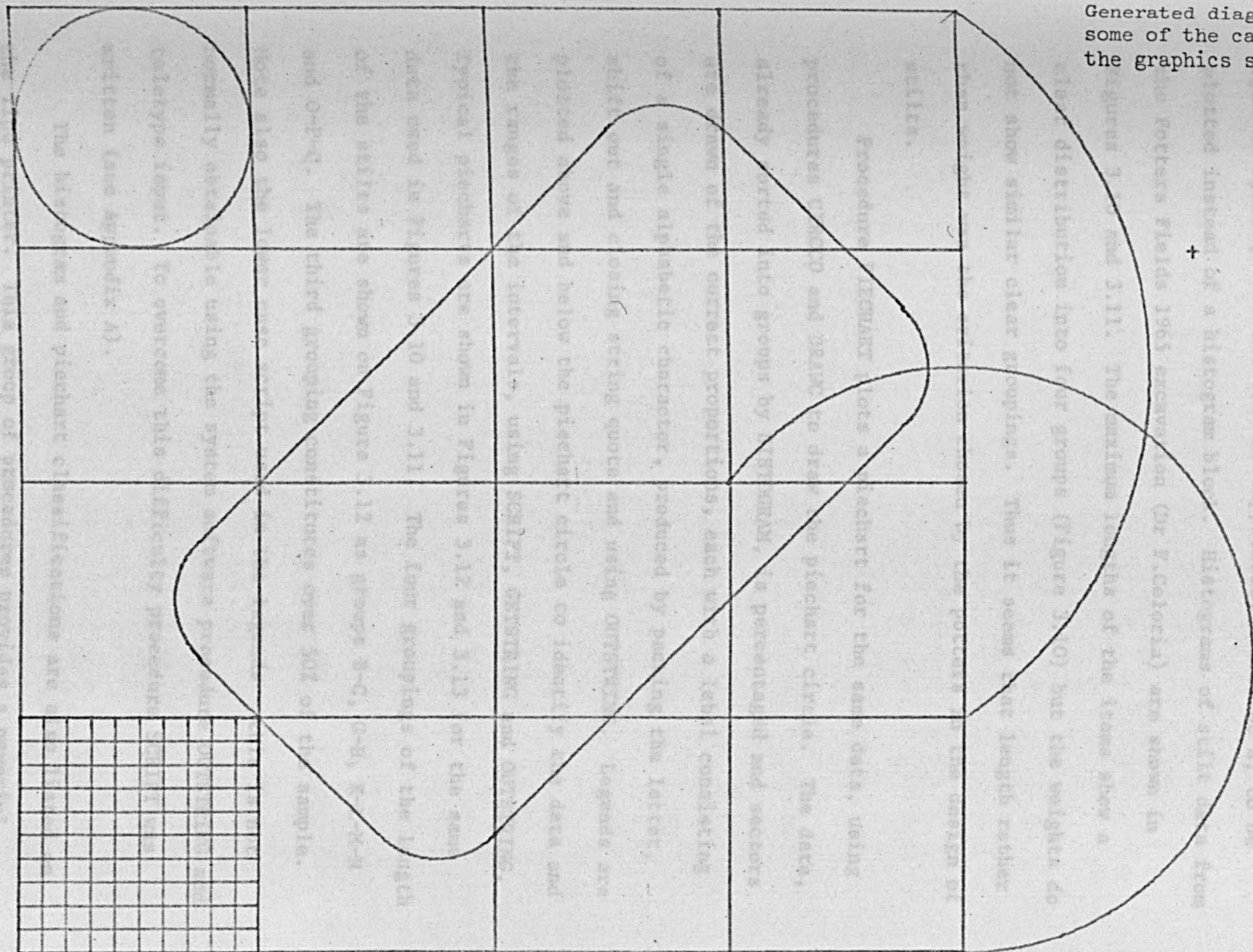
Plotting commands for
numerical items

Specification and drawing
of arcs

Change origin

Plotting symbol output

Generated diagram showing
some of the capabilities of
the graphics system



SITE GRID AND ROMAN CAMP

Figure 3.9

pollen analysis, figure 6.1).

"Trace" readings, i.e.

non-zero readings less than a specified small value cause a special symbol, + if the axis is suppressed, otherwise X, to be plotted instead of a histogram block. Histograms of stilt data from the Potters Fields 1965 excavation (Dr F.Celoria) are shown in Figures 3.10 and 3.11. The maximum lengths of the items show a clear distribution into four groups (Figure 3.10) but the weights do not show similar clear groupings. Thus it seems that length rather than weight was the criterion chosen by the potters in the design of stilts.

Procedure PIECHART plots a piechart for the same data, using procedures CIRCCO and DRAWC to draw the piechart circle. The data, already sorted into groups by HISTOGRAM, is percentaged and sectors are drawn of the correct proportions, each with a label consisting of a single alphabetic character, produced by packing the letter, shift out and closing string quote and using OUTSTRING. Legends are plotted above and below the piechart circle to identify the data and the ranges of the intervals, using SCRIPT, GETSTRING and OUTSTRING. Typical piecharts are shown in Figures 3.12 and 3.13 for the same data used in Figures 3.10 and 3.11. The four groupings of the length of the stilts are shown on Figure 3.12 as groups B-C, G-H, K-L-M-N and O-P-Q. The third grouping constitutes over 50% of the sample. Note also the lower case script used in the legends - this is not normally obtainable using the system software procedure OUTSTRING and teletype input. To overcome this difficulty procedure SCRIPT was written (see Appendix A).

The histogram and piechart classifications are also listed on the line printer. This group of procedures provides a powerful facility for the archaeologist, who can submit his raw data for

POTTERS FIELDS 65 STILT ANALYSIS

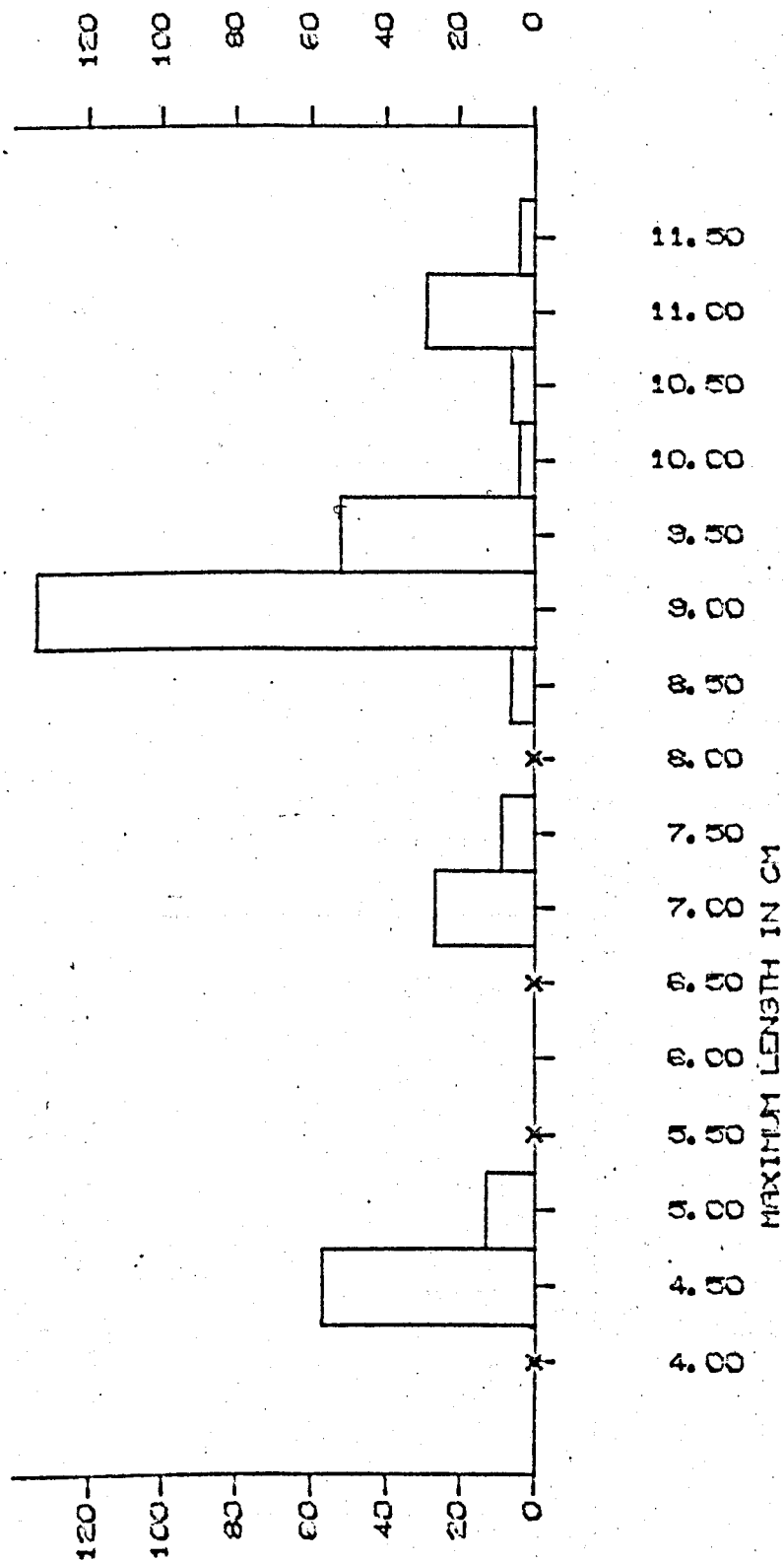


Figure 3.10

POTTERS FIELDS 65 STILT ANALYSIS

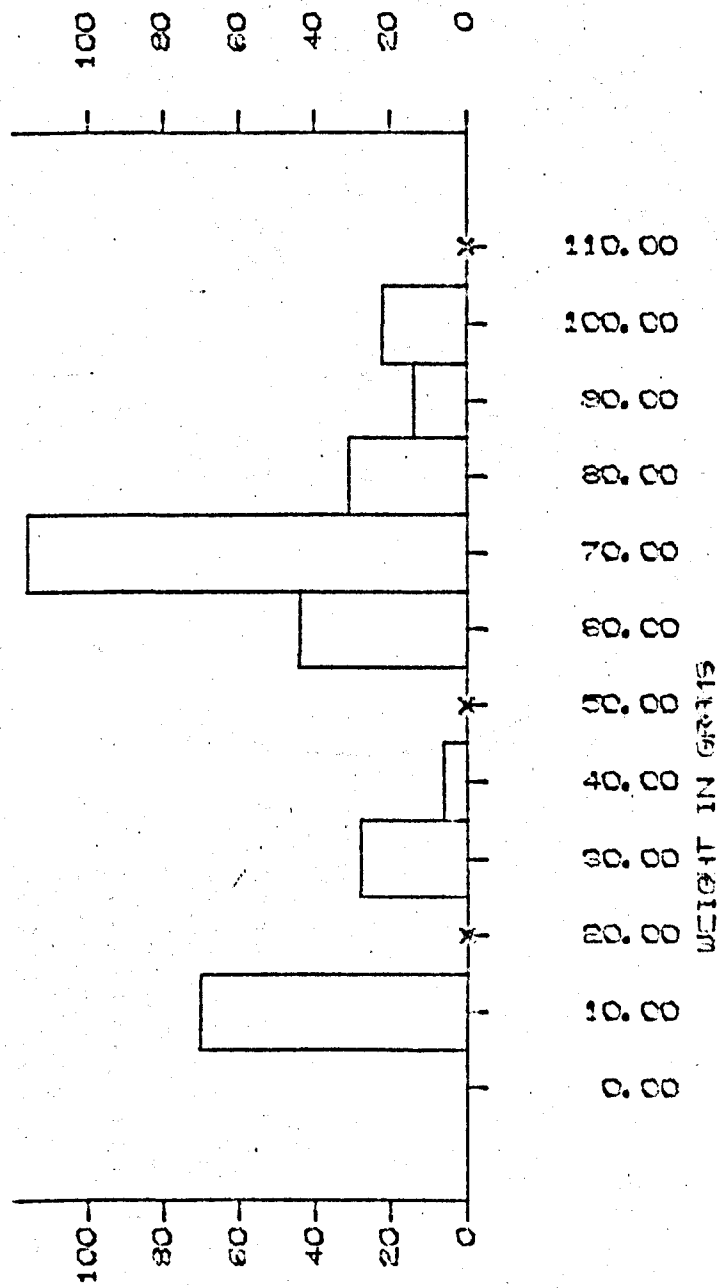
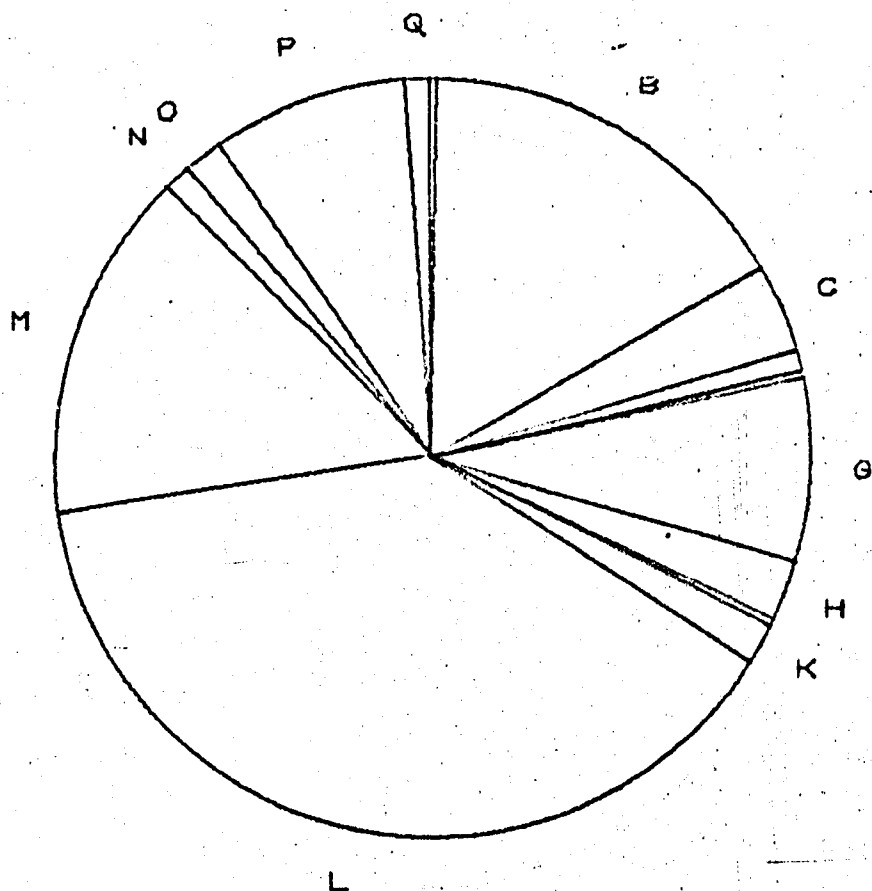


Figure 3.11

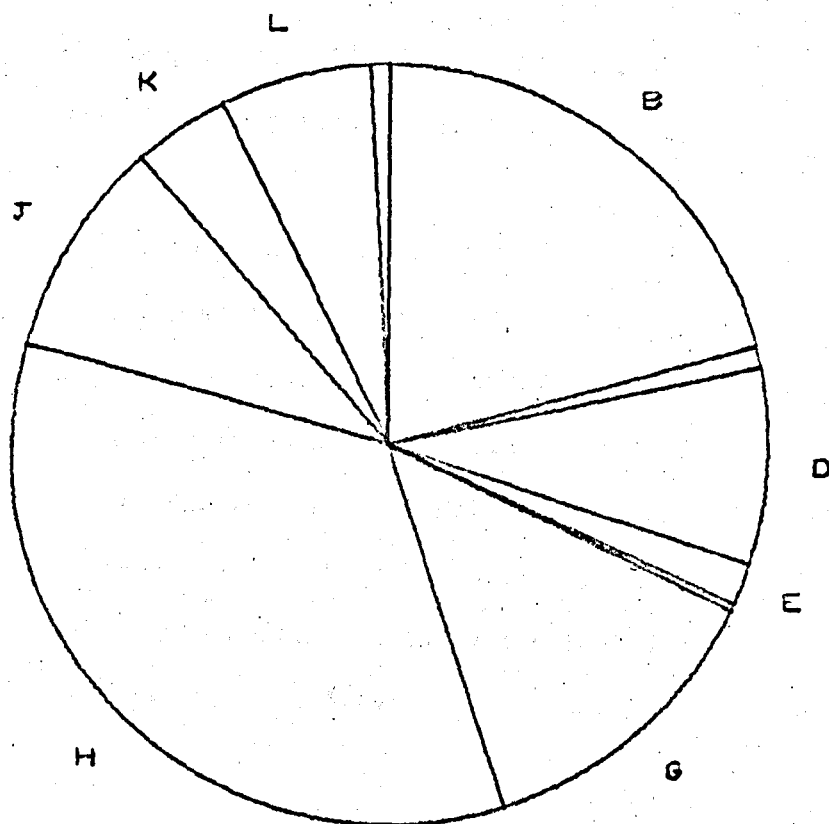
Piechart
of POTTERS FIELDS 65
Stilt Analysis
by Maximum Length



A	Maximum Length	3.75 - 4.24 cm
B	4.25 - 4.74 cm	
C	4.75 - 5.24 cm	
D	5.25 - 5.74 cm	
E	5.75 - 6.24 cm	
F	6.25 - 6.74 cm	
G	6.75 - 7.24 cm	
H	7.25 - 7.74 cm	
I	7.75 - 8.24 cm	
J	8.25 - 8.74 cm	
K	8.75 - 9.24 cm	
L	9.25 - 9.74 cm	
M	9.75 - 10.24 cm	
N	10.25 - 10.74 cm	
O	10.75 - 11.24 cm	
P	11.25 - 11.74 cm	
Q		

Figure 3.12

Piechart of POTTERS FIELDS 65 Stilt Analysis by Weight



A	Weight less than 4.99 gm
B	5.00 - 14.99 gm
C	15.00 - 24.99 gm
D	25.00 - 34.99 gm
E	35.00 - 44.99 gm
F	45.00 - 54.99 gm
G	55.00 - 64.99 gm
H	65.00 - 74.99 gm
I	75.00 - 84.99 gm
J	85.00 - 94.99 gm
K	95.00 - 104.99 gm
L	105.00 - 114.99 gm

Figure 3.13

histogram and piechart analysis without any previous specification of intervals, ranges or number of readings.

3.4.8

ARCJW015 Pollen analysis (ALGOL 60)

This program was developed in collaboration with a pollen analyst, Miss K.Simpkins, of the University of Keele. Pollen analysis is used by archaeologists to determine climate in previous ages and the "*Ulmus decline*" together with an increase in proportion of agricultural weed pollens indicates the onset of the Neolithic farming revolution. The program uses procedure HISTOGRAM to plot pollen diagrams. It is fully described in paragraph 6.12.

3.4.9

ARCJW016 Map magnetic tape copy and additions (ALGOL 60)

This program amends a map master tape as used by programs ARCJW011 (see paragraphs 2.6.6 and 3.4.5) and ARCJW022 (see paragraph 3.4.14). The old and new master tapes are initialised. This is a tailor-made program which standardised the old MAPGBI and MAPSTAFFS maps, writing the standardised maps to the new master tape. It uses procedures OUTSBL and SEARCH and is now fully superseded by READMAP (see Appendix A and paragraph 3.4.14).

3.4.10

ARCJW017 MAPSTAFFS test program (ALGOL 60)

This is a skeleton program to test routines MAPGBI and MAPSTAFFS in conjunction. Each of these routines plots a specific map, reading the coordinates from magnetic tape i.e. from the map master tape created by ARCJW016 (paragraph 3.4.9 above). The start of the map coordinates is indicated by a special block consisting of the characters M A P succeeded by the specific identifier. All maps are given borders, grid north points and scales, and grid references may be plotted to give a distribution map. Each routine is, however,

tailor-made and both are now superceded by DRAWMAP and PLOTREFS (see Appendix A and paragraph 3.4.14).

3.4.11

ARCJW019 Pottery profile suite (ALGOL 60)

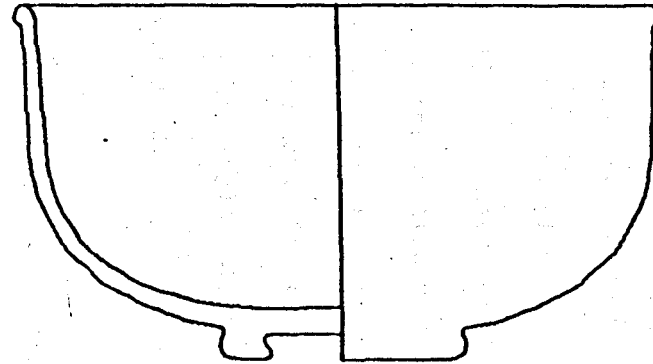
This program reads coordinates of pottery profiles, normalises them and calculates profile parameters. In the graphics mode it also draws diagrams of the pots in the conventional archaeological manner, as illustrated in Figure 3.14. The program is fully described in paragraph 4.4.3.

3.4.12

ARCJW020 New font test program (ALGOL 60)

This program develops the use of the system software routine INSTRING to define a sequence of upper and lower case characters, numerals and punctuation marks, which may then be output in any type face designed by the user. Thus by the use of this facility the user is no longer confined to system typeface. The example specified is the Optical Character Recognition font OCRB proposed by the International Standards Organisation, and a slightly simplified version similar to that designed by the National Physical Laboratory (Heap and Laws, 1968) has actually been implemented. The NPL version was in KDF9 Usercode while the version in this thesis is defined using ALGOL 60. The routine makes frequent use of the procedure O(A,B,C,D) which converts the octal parameters A,B,C and D into binary and then packs all four parameters into a single 24-bit word for economy of storage. The packed string produced by INSTRING is unpacked by a (NEAT) machine code sequence then examined character by character. Considering each character as octal, 00-47 specify coordinates on a 4 x 7 grid, the first digit giving the x displacement and the second the y displacement.

SAMIAN DRAG. 37



SAMIAN DRAG. 29

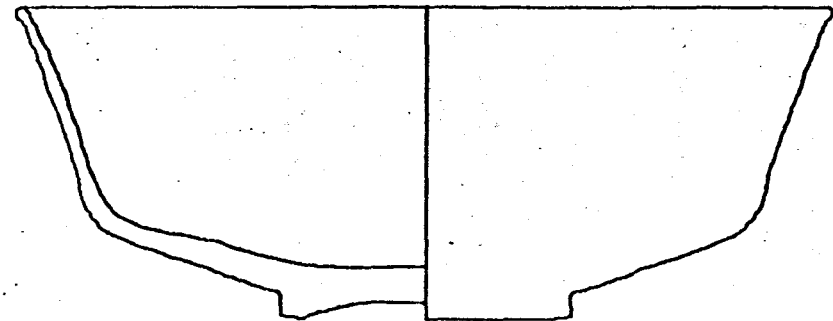


Figure 3.14

Samian pottery diagrams

5Y causes the value Y to be added to y displacements until further notice, thus raising the character above the normal position, while 6Y similarly lowers the character by value Y. 7Y causes the pen to be raised, and 77 indicates the end of a character. The next character is drawn 5 units to the right of the last. An italic feature is available, the x displacement being progressively increased as the y displacement increases, and the degree of the slope of italicised characters depends on the value of the italic parameter. An orientation parameter and a size parameter control these features of the typeface in a similar manner to the system software routine WAY. The coding for OCRB is given in Appendix A, Figure A1, while sample outputs of the user-designed typeface are shown in Figure A2.

3.4.13

ARCJW021 Masons' marks analysis (ALGOL 60)

This program analyses masons' marks, recording their significant features. A by-product is a plot of the marks in normalised form (see Figure 1.7). During the pattern recognition process the marks are compared by rotation, translation and scaling, and this process may also be plotted (see Figure 1.8). The program is fully described in paragraph 1.10.3.

3.4.14

ARCJW022 General map generator (ALGOL 60)

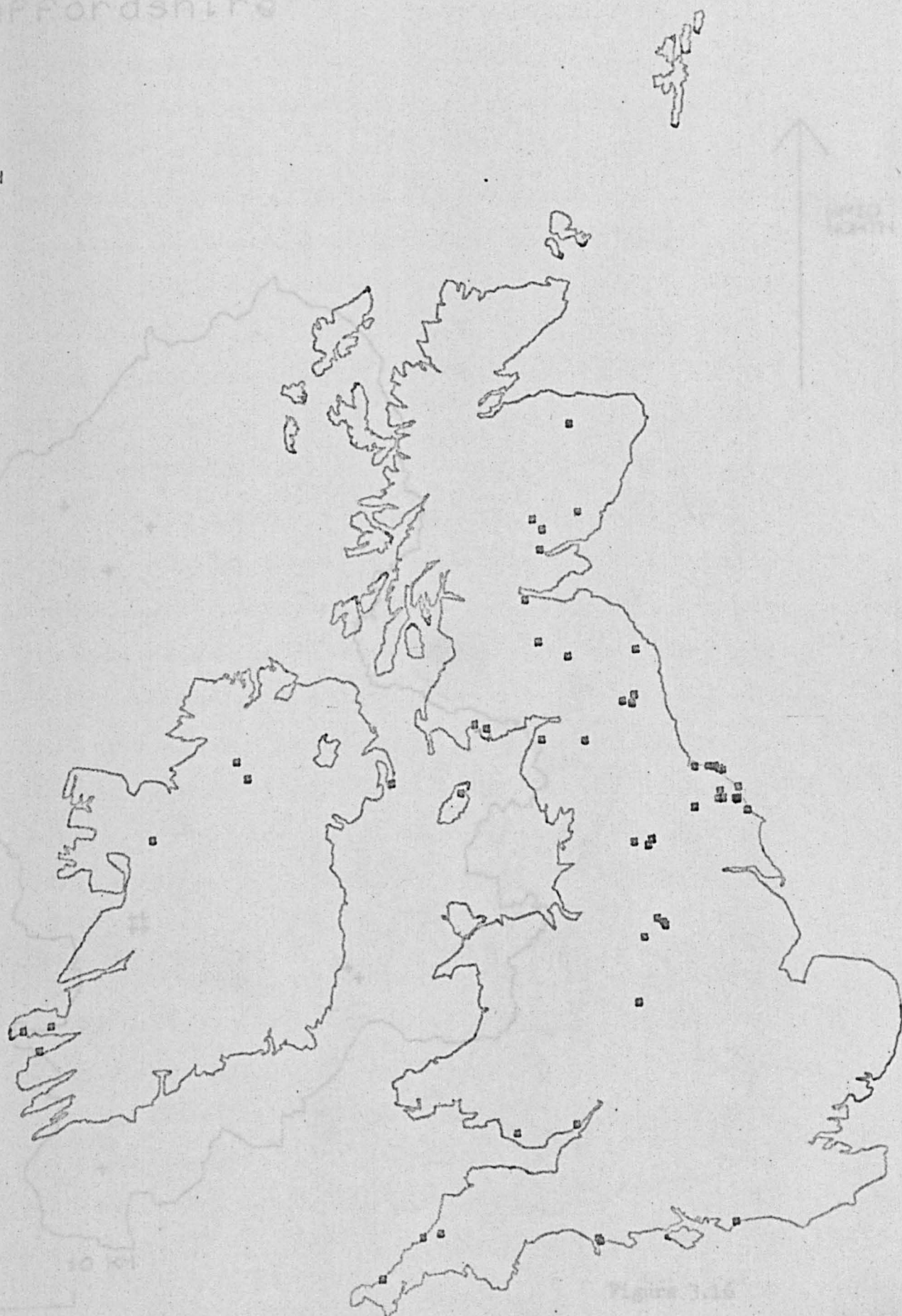
This program is similar in many aspects to ARCJW011 (see paragraphs 2.6.6 and 3.4.5). The information retrieval aspect is, however, absent, and all the data specifying the map, titles and grid references to be plotted are read from paper tape. The main program is a skeleton which calls DRAWMAP and then PLOTREFS. For completion in the generalisation process evolved from the custom-built programs

used to generate, draw and plot references on MAPGBI and MAPSTAFFS, a procedure READMAP is included. This procedure reads from the paper tape reader the following information: the character identifier of a map, terminated by *; the scale of the original map; the scale of the desired reproduction; d-Mac coordinates for the grid origin and a point grid north of the origin, followed by the *halt* code; a complete set of data for each detached line or closed curve on the map, as follows: for other than the first line a non-zero integer (a zero value in this place indicates that the map is complete); offsets in Km east and north for the sub-origin of this line relative to the origin (this is to allow map inserts to be traced); d-Mac coordinates for the sub-origin and up to 3000 desired points on the line, followed by the *halt* code. If offsets are specified as zero, the origin is taken as the sub-origin and the origin coordinates must be specified in place of the sub-origin coordinates. The procedure rotates the map to a truly vertical alignment, x axis to grid north, y axis to grid west, translates all coordinates to a zero origin and scales as required. Finally, the coordinates are written to the magnetic tape new master, which has originally been copied from the old master at the beginning of the procedure. The data is preceded by map identifier and scale. Another map is then processed (if no further map is present, * is used as terminator). The program uses the following procedures: READMAP, DRAWMAP, PLOTREFS, GETSTRING, SCRIPT, OUTSBL and SEARCH.

DRAWMAP and PLOTREFS are described in full in Appendix A. Their versatility is illustrated by Figures 3.15 - 3.17. Note that the scale of the map may be varied at will; the north points are positioned automatically in different places on different runs so that overwriting of part of the map outline is impossible; the scale at the bottom left hand corner is adjusted as necessary; and the border is at least A4 size (Figures 3.16 and 3.17) but may be larger to accommodate legends

UNCHAMBERED CAIRNS AND BARROWS
WITH DECORATIVE CARVINGS
OF ANY CLASS
TYPE B (BARROW) SITES
WITH DECORATIVE CLASSES 1-8)

GRID
NORTH



100 101

Figure 3.15

Distribution
of Roman Sites
in the County of
Staffordshire

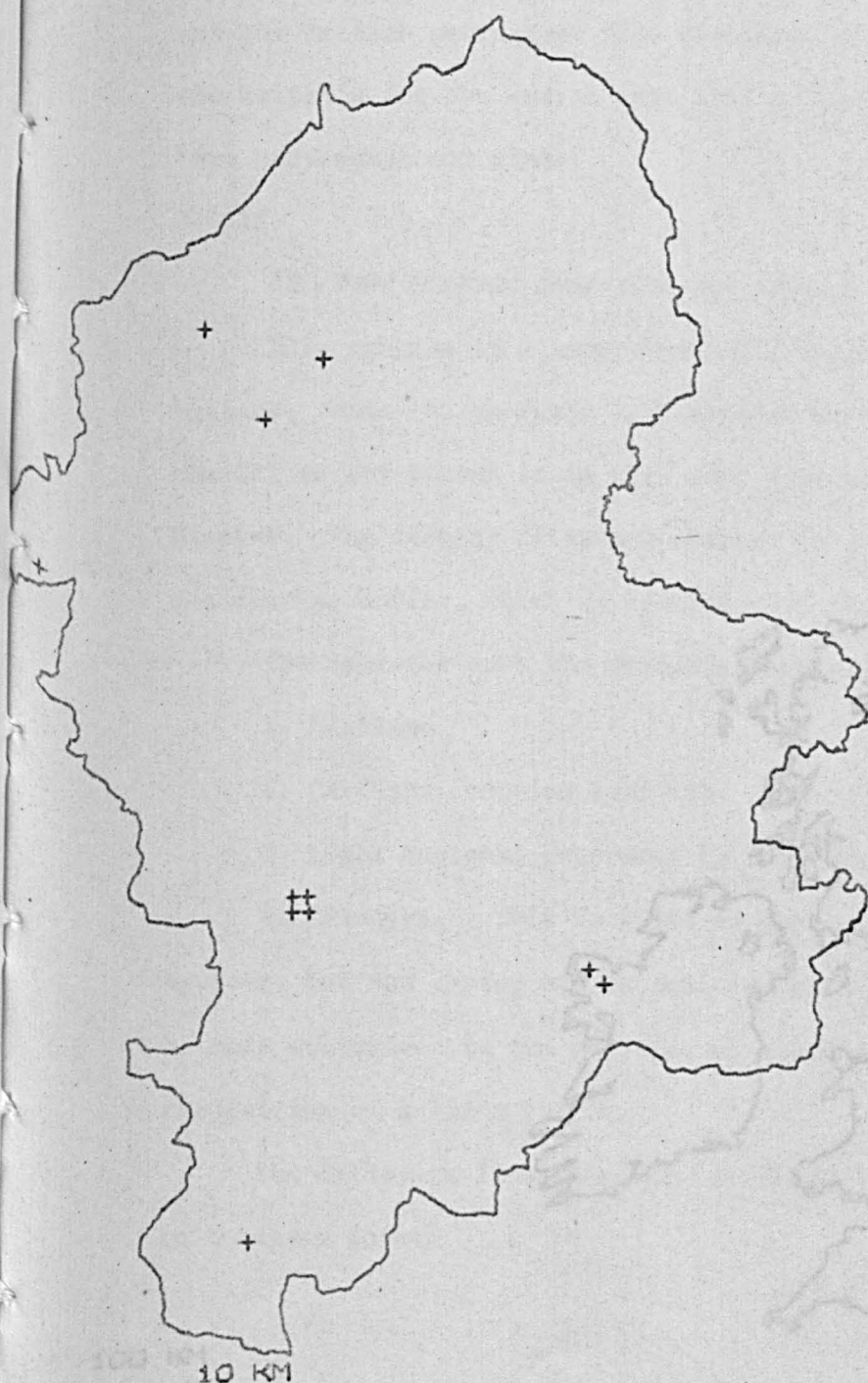
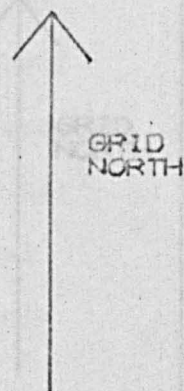


Figure 3.16

Roman Legionary Bases in Britain

...for the plotter where available. The map should be automatically truncated where it hits the border. Figure 3.15 should be compared with Figure 3.1, and Figure 3.16 with Figure 3.2. The differences between ARJW024 and ARJW025 will now be noted. Note also that Figure 3.15 is the output of an ARJW024 program run using DRAWMAP and PLOTTER, i.e. the ARJW024 program and its routines, as illustrated by the program. The program was the British petroglyph file consisting of the criteria for the search and the criteria for the selection of petroglyphic carvings of any class.

3.4.15

ARJW024 Diagram generator for the British petroglyphic file

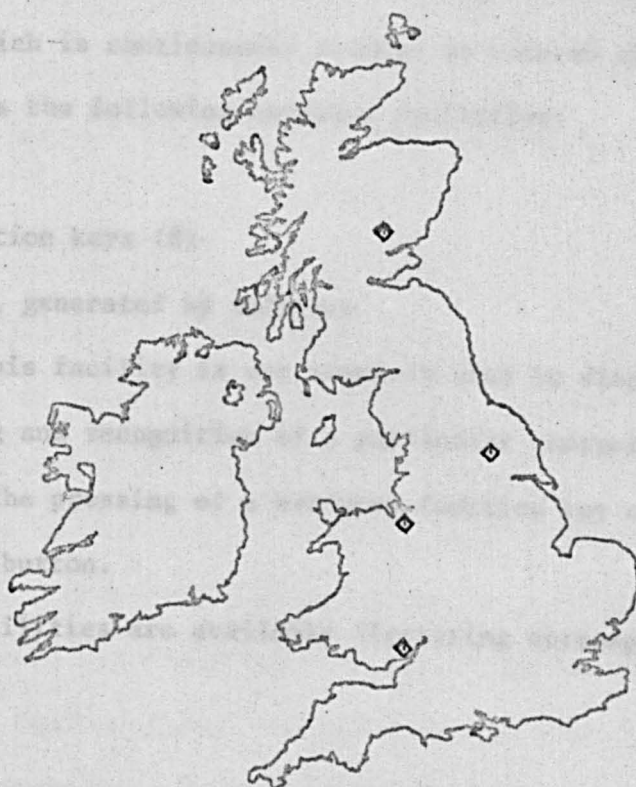
This program is a conversational program for the petroglyphic file. When the operator is satisfied with the results of the search on the CRT he may record it to tape or to a plotter. The display files are written to the display buffer, which is available to the operator.

The operator uses the following:

1. Lightpen
2. Hardware function keys (F1-F10)
3. Light buttons, generated by the program
4. Teletype. This facility is available for the teletype systems, but the typing and the teletype input can be made equivalent to the light pen input by the use of the recognition of a light pen.

The following facilities are available to the operator (in addition to teletype input):

The following facilities are available to the operator (in addition to teletype input):



100 KM

Figure 3.17

as necessary (Figure 3.15). If the scale of the map specified is too large for the plotter width available, the map outline is automatically truncated where it hits the border. Figure 3.15 should be compared with Figure 2.7, and Figure 3.16 with Figure 2.8. The differences between ARCJW022 and ARCJW011 will then be apparent. Note also that Figure 3.15 is the output of an information retrieval run using DRAWMAP and PLOTREFS, i.e. ARCJW011 modified with the general routines, as illustrated in the program listings. The data searched was the British petroglyph file described in paragraph 2.6.2.2 and the criteria for the search were type B (barrow) sites with decorative carvings of any class.

3.4.15

ARCJW024 Diagram generator for display unit (ALGOL 60 and EDGAR)

This program is a comprehensive diagram generator for archaeologists. When the operator is satisfied with a diagram displayed on the CRT he may record it in hard copy form on the digital incremental plotter. The display files are carried on disc and generate code in the display buffer, which is continuously scanned to refresh the tube.

The operator uses the following hardware facilities:

1. Lightpen
2. Hardware function keys (8)
3. Light buttons, generated by software
4. Teletype. This facility is not commonly used by display systems, but the typing and recognition of a particular character may be made equivalent to the pressing of a hardware function key or recognition of a light button.

The following facilities are available (lettering corresponds to teletype input)

- A) draw International A proportions frame, vertical orientation.
- B) alter Brightness level (initially set at medium brightness).

Three levels are available, dim, medium and bright. This option applies to the next feature only.

- C) specify Coordinates of a point on the display explicitly.

A coordinate frame is temporarily displayed. The operator then specifies whether the coordinates are to describe a single point, the end point of a new line to be drawn, or the position of text. The X and Y coordinates are then specified explicitly via the teletype (0-1023). The program then enters the point specification, line generation or text generation routines.

- D) Delete whole diagram.

- F) set Frame marker. This causes all lines generated subsequently to be truly horizontal or vertical. The position of the light pen indicates only one coordinate of the endpoint of the new line, the other being determined by the horizontal or vertical property. If a predefined point is indicated as the endpoint of the new line the previous line and the new line are adjusted to give a true right angle. (This is an unusual feature in a graphics program).

- H) Horizontal A proportions frame.

- M) Multidimensional scaling plot. The input is a paper tape which specifies the type of symbols to be plotted (+, diamond, X, dot), the coordinates of a 2D MDSCAL output and the labels to be attached to the points. The coordinates are scaled without distortion to fit the vertical A proportions frame on the screen, plotted and labelled. Several different types of symbol may be used on the same diagram. Renfrew-Sterud linkage (probably the first conjunction of these techniques) may be added using the line generator, and isopleths may be drawn around clusters using the sketching

facility. These last two features improve the information given on the original MDSCAL plot.

- N) Reset frame marker (draw Normal lines subsequently). The position of the light pen indicates both coordinates of the endpoint of a new line.
- O) draw Circle with centre at the last specified point and of radius specified on the teletype.
- P) Phenon (dendrogram) plot. The input is a paper tape which specifies the number of items in the dendrogram, the reference numbers of the items in their desired order, labels for the items, and groups of three numbers, each specifying the percentage level of the next join, and the two reference numbers of the items to be joined. The process continues until all items have joined into one group. A percentage scale is drawn at the left-hand side, and items are labelled at their head. The whole diagram fits the horizontal A proportions frame. Two or more items may join at any given percentage level.
- S) Sketch. The routine switches the tracking cross on, then waits for a single hardware function key to be pressed. This is the signal that the lightpen is positioned at the start of the desired curve. Thereafter the routine waits until the modulus of the x change or modulus of the y change exceeds a specified value, then draws a small straight-line vector to the new position, adding the vector to the old buffer. Next the routine waits for a specified number of frame holds, then deletes the curve, returning to the point where the moduli of the x and y changes are checked. The curve continues to be extended until a single hardware function key is pressed before the specified number of frame holds occur. This is taken as

the signal for the end of the curve. It is believed that some novel features have been provided in this algorithm, the essentials of which are as follows:

```
NEWBUF(BUF); SETTRAK(true);
FOUND:=0;
WAIT: I:=ACTION(1,FOUND); if FOUND#1 then goto WAIT;
comment await pressing of function key signalling correct
positioning of lightpen at start of curve;
PENTRAK(P1X,P1Y);
POINT(P1X,P1Y,0); SETP(BR,CH,0,0);
comment record start of curve and set parameters;
SLOOP: PENTRAK(P2X,P2Y);
if ABS(P2X-P1X)>2 or ABS(P2Y-P1Y)>2 then begin
OLDBUF(BUF);
VECTOR(P2X-P1X,P2Y-P1Y,1); P1X:=P2X; P1Y:=P2Y; end else goto SLOOP;
comment if modulus of x change or modulus of y change exceeds
specified value then add vector to buffer;
DRAW(NAME,0);
FOUND:=FH;
comment set up desired number of frame holds to occur in
default of function key being pressed signalling end of
sketched curve;
I:=ACTION(3,FOUND);
if FOUND#2 then goto EXIT;
DELETE(NAME); goto SLOOP;
```

In practice the modulus change in x or y and the number of frame holds may be adjusted to take account of the coarseness of the required curve and the size of the current display file (which affects the refresh rate and consequently the frame hold delay).

T) Switch Tracking cross on

U) switch tracking cross off (Untrack). This is necessary for the final plotting of the diagram from the display file if the tracking cross is not to appear in the hard copy.

Options using hardware function keys (numbering corresponds to number of function key). These are frequently-used options.

- 1) Specify point. The coordinates of the tracking cross are taken as the defined point, which is added to the point list.
- 2) Retrieve coordinates of previously defined point. The numerical identifiers of all points in the point list are displayed in their correct positions by PTNUMS. The desired point may be indicated by using the lightpen to "see" the corresponding identifier or by typing the identifier on the teletype (this option is useful if the identifier is obscured by other features).
- 3) Draw line from last specified point to current position of tracking cross. If the frame marker is reset the line is drawn direct, otherwise a truly horizontal or vertical line is drawn depending on the relative magnitudes of the x and y changes (in this case only the x or only the y coordinate being determined by the previous point). The end of the new line is added to the point list and this point then becomes the current point.
- 4) Draw line from last specified point to previously defined point. The numerical identifiers of all points in the point list are displayed in their correct positions by PTNUMS.

The desired termination point for the new line may be indicated by using the lightpen to "see" the corresponding identifier or by typing the identifier on the teletype (this option is useful if the identifier is obscured by other features). If the frame marker is reset the line is drawn direct, otherwise a truly horizontal or vertical line is drawn depending on the relative magnitudes of the x and y changes (this involves repositioning the previous point and redrawing the previous line: the coordinates of the previous point are also altered in the point list). The end of the new line becomes the current point, but since it is already in the point list it is not added to this list.

- 5) Delete item indicated by lightpen. If this is the last specified item a marker is set to indicate that no modifications can be carried out on the last specified item (see function key 7).
- 6) Defines a point, either a new point by means of the tracking cross or a previously-defined point using PTNUMS. This is used to position text on the screen. The brightness and character size are defined by teletype messages, then the actual text to be displayed is typed in and terminated by full stop.
- 7) Modify last specified item by use of light buttons, which are displayed along the bottom of the screen. The following options are available:

MOVE) Move item to position of last defined point - it is necessary to define this point prior to calling the light buttons by hardware function key 7.

ROT) Rotate item 0.01 radians about last defined point in an anticlockwise direction.

SCUP) Scale item up by factor of 1%

SCDWN) Scale item down by factor of 1%

STOP) Remove light buttons and await next action.

MOVE similarly causes the removal of the light buttons, but ROT, SCUP and SCDWN retain the light buttons, one cycle of the modification taking place each time the light pen "sees" the corresponding light button and modification takes place continuously while the light pen is held at the light button. Combinations of ROT, SCUP and SCDWN may thus be executed in sequence.

- 8) Read pot profile specified by d-Mac coordinates. Uses READFIG. The pot is drawn in conventional left-hand section and right-hand elevation with suppression of re-entrant footrim. Vertical centre line and horizontal rim line are added. DDRAWF is used to draw the profiles. A title is added, using UNPACK and OUTPUT, and the scale of the drawing. The first drawing is actual size (scale = 1.0) but this may be altered to any scale (e.g. $\frac{1}{4}$ scale = 0.25) by typing 100 * scale on the teletype and using DECIN (thus for $\frac{1}{4}$ scale, 25 is typed). The modification cycle continues until the operator indicates his satisfaction by typing 0 in response to the SCALE: message.

The program uses the following procedures: UNPACK, READFIG, DDRAWF (Display draws figure. Expects string of coordinates with termination codes as created by READFIG), PTNUMS (displays all point identifiers at bright illumination and exits when one identifier is indicated by lightpen "hit" or by typing it on the teletype), DECIN. Using this facility, the archaeologist can generate diagrams which are ready for publication.

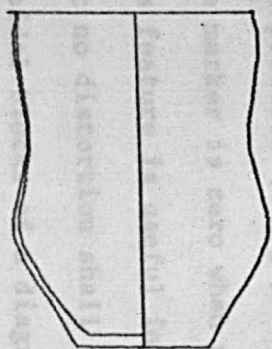
The versatility of this program may be judged from the following Figures which have been generated by it:

Figures 1.1, 1.3, 1.5 (Scalograms)

Figures 1.2, 1.4, 1.6, 1.18, 1.20, 1.21 (Dendrograms)

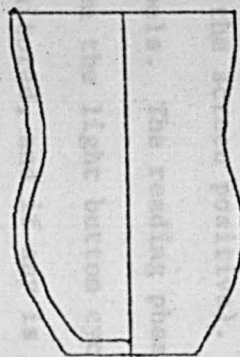
Figure 3.18 (British Beaker pottery conventional profiles with scales in A4 border).

BRITISH BEAKER POTTERY - CLARKE CLASSIFICATION



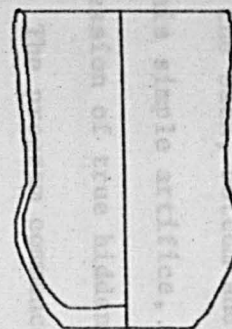
BEAKER S1 751

SCALE 0.25



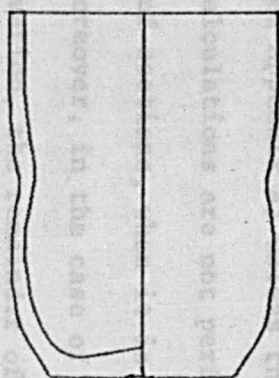
BEAKER S2W 807

SCALE 0.25



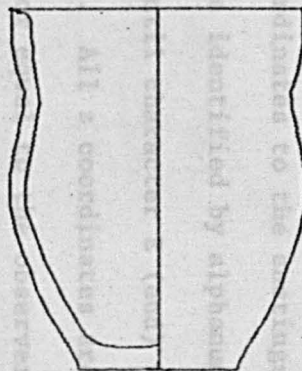
BEAKER S2E 846

SCALE 0.25



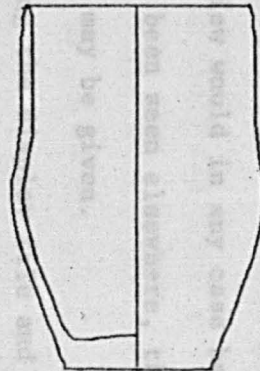
BEAKER S3E 935

SCALE 0.25



BEAKER N4 717

SCALE 0.25



BEAKER S4 977

SCALE 0.25

Figure 3.18

3.4.16

ARCJWO25 3D rotation and perspective (ALGOL 60 and EDGAR)

This program rotates a diagram in three dimensions, displaying a perspective view on the screen. The "hidden line" problem is avoided in the case of archaeological block diagrams by displaying the front, top and one side view only and rotating the figure so that the back, bottom and other side view would in any case be hidden. By this simple artifice, which has not been seen elsewhere, the impression of true hidden line removal may be given.

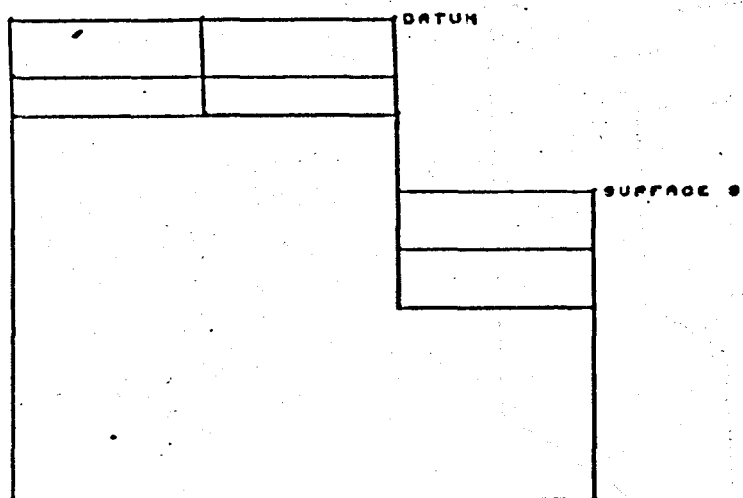
The program commences by resetting the display file and initialising parameters. A series of line (prefix L) and point (prefix P) specifications are read from paper tape, all coordinates being expressed in northings, eastings and depths from the origin. The screen x coordinates are set to the northings, y coordinates to the negatives of the depths, and z coordinates to the eastings (out of the screen positive). Points may be identified by alphanumeric labels. The reading phase continues until character E (end) is read, when the light button cycle is entered. All z coordinates are monitored, and if any is greater than or equal to the observer's position (which is initialised at 30 inches from the screen for perspective calculations), the observer's position is moved further out by $\frac{1}{2}$ inch, i.e. the observer is not allowed to move inside the displayed object. Each point is displayed in its simulated perspective position (using procedure PERS, see Appendix A) unless the perspective marker is zero when perspective calculations are not performed - this feature is useful in the display of sections, when it is desired that no distortion shall take place; moreover, in the case of the cuboidal blocks of a diagrammatic excavation, the remainder of the sides of the block are hidden behind the boundary lines of the section

itself and are not apparent in the section displayed. This feature is illustrated in Figure 3.19.

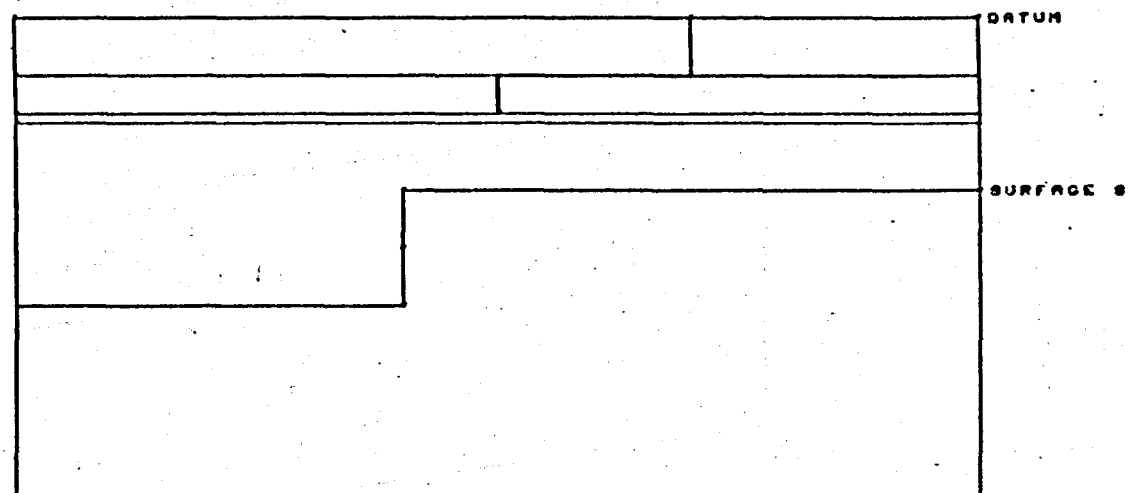
Labels for the points are added in small characters. Lines are drawn from the previous coordinates to the current coordinates, each in perspective (unless the perspective marker is zero) and the illusion of depth is enhanced by brightening lines which are near the observer and dimming those which are far away. The lines are drawn at dim, normal or bright illumination depending on whether the line is wholly behind, intersects or is wholly in front of the screen plane. A series of light buttons is displayed as follows (see Figure 3.20):

XYROT	Rotate about the Z axis 5° .
YZROT	Rotate about the X axis 5° .
ZXROT	Rotate about the Y axis 5° .
ZMIN	Zoom observer in.
ZMOUT	Zoom observer out.
PNRT	Pan view to right (object moves left).
PNLFT	Pan view to left (object moves right).
ZX+90	Rotate about the Y axis 90° .
STOP	Remove the light buttons so that hard copy may be taken, and await a function key depression.

A wait loop is entered where a light pen "hit" is awaited, the light button selected determining the action taken, as above. The modified figure is displayed, and the cycle repeats. During the normal light button cycle the light buttons are not continually written into the display file, as is the modified figure. The STOP and ZX+90 light buttons cause the light buttons to be deleted, and a wait loop is entered where the depression of a function key is awaited, as follows:



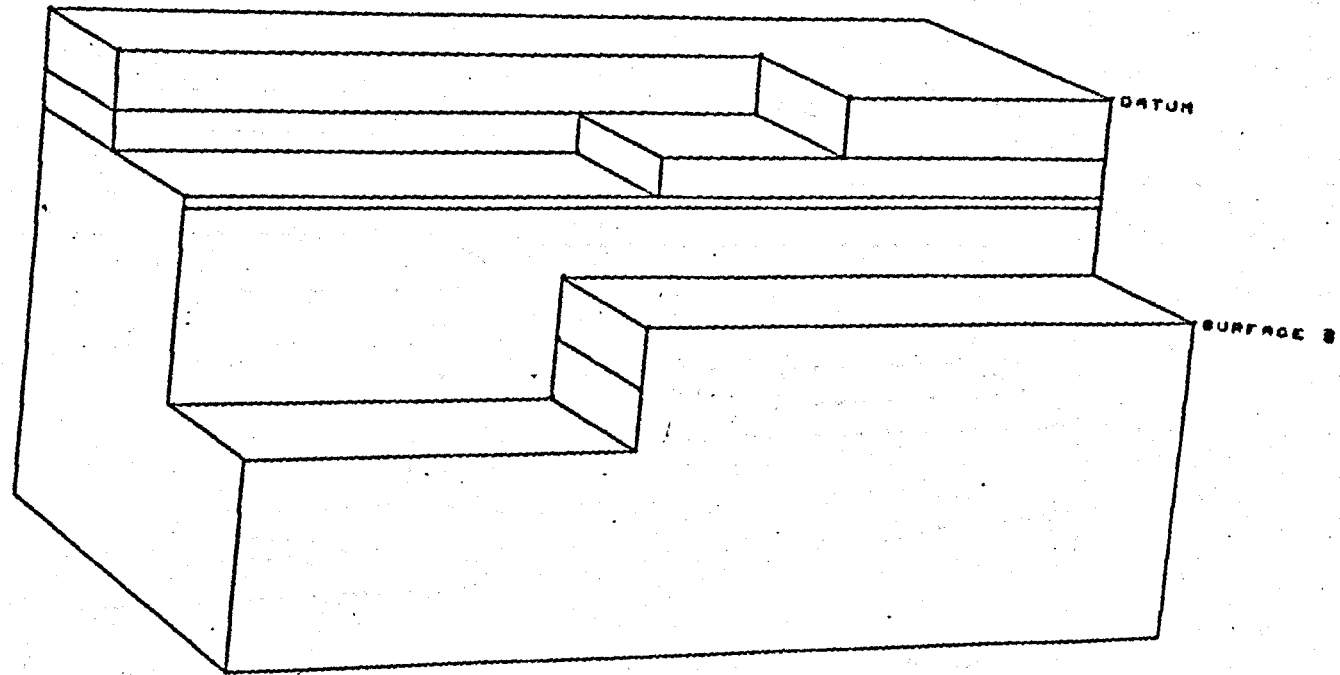
Left-hand section



Front section

Two different sections without perspective of the same archaeological block diagram

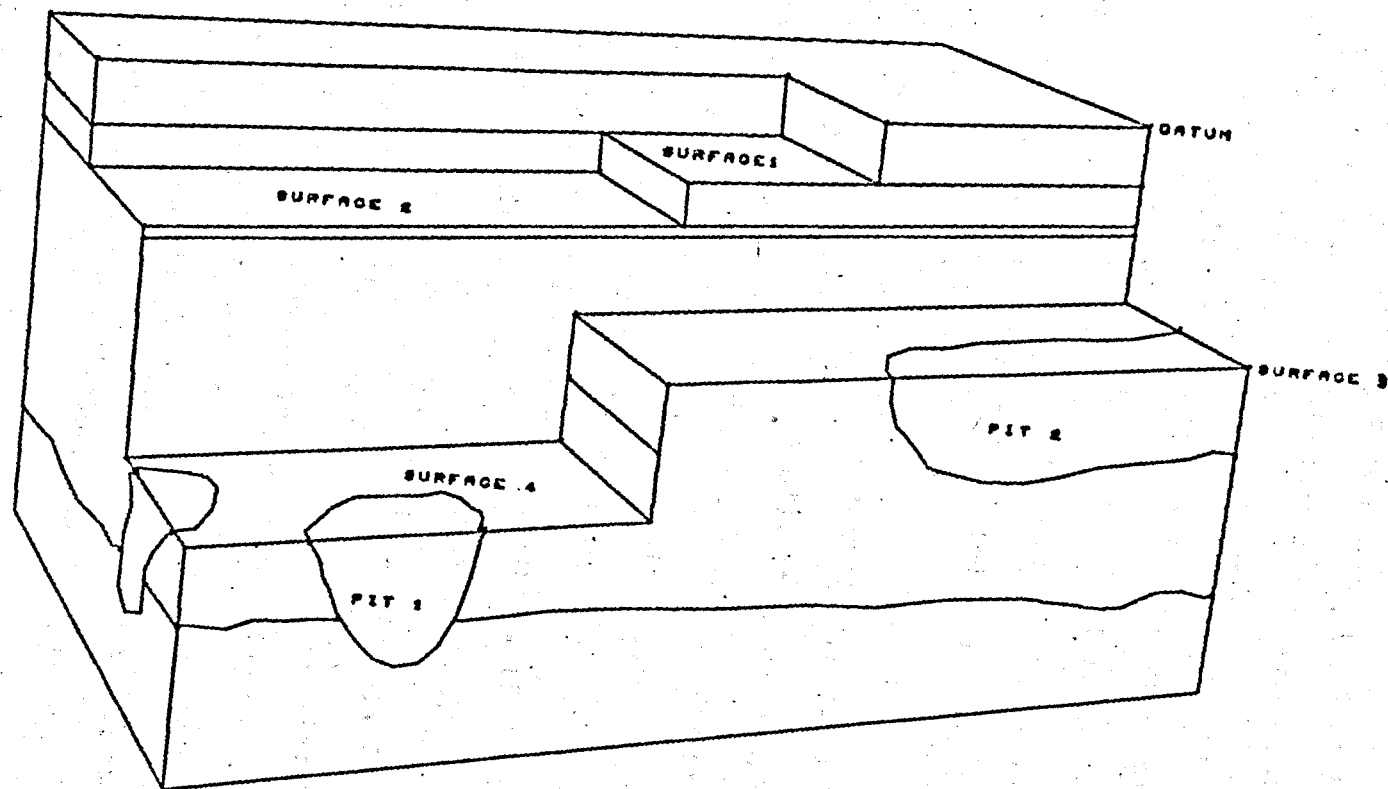
Figure 3.19



XYROT YZROT ZXROT ZMIN ZMOUT PNRT PNLFT ZX+90 STOP

Perspective block diagram, showing light buttons

Figure 3.20



Perspective block diagram with sketch and text facilities

Figure 3.21

Function key 1. Enter the light button cycle again

2. Delete the whole display and restart the program

3. Sketch a freehand curve using the lightpen - this is the same routine as is used in ARCJW024, with a different number of frame holds to allow for the differing lengths of the display files

4. Alter the sense of the perspective marker, i.e. turns the perspective off if it is on, and vice versa, modifying the displayed figure accordingly

5. Delete the item indicated by the lightpen "hit"

6. Insert text.

Since the lightpen tracking routine is not normally operative (it is switched on specially for sketching and text), it is necessary to enter a waiting loop whenever keys 3 or 6 are pressed. The tracking cross is then moved to the desired starting position using the lightpen, and this is indicated by pressing any function key. In the sketching option the tracking routine remains active, sketching a line until a function key is pressed again, when the tracking cross is removed. In the text option the tracking cross is removed immediately, text is entered via the teletype and displayed in the smallest character size. Function keys 3,5 and 6 keep the program in the key-setting mode. Function key 4 alters the perspective marker then enters the light button mode. Function key 1 enters the light button mode immediately, while function key 2 deletes the whole display, reads a new figure then enters the light button mode.

This program is very useful to the archaeologist in the construction of 3D perspective block diagrams of excavations. The sketching facility may be used to add pit outlines, etc. and the text facility to

add labels (see Figure 3.21). The rotations, panning and zooming are used to get a presentable figure, switching the perspective off without rotation gives a front section, while ZX+90 with perspective off gives a side section. The program uses procedures ROT (for rotation) and PERS (for perspective calculations), which are described in Appendix A.

3.4.17

ARCJW026 Pit scanner (ALGOL 60 and EDGAR)

This program is useful in the publication of pit diagrams. The routine resets the display file, then sets the observer's position at 30 inches from the screen plane. The specification of the 3D outline of an excavated pit is read from paper tape. First the maximum depth of the pit is specified, then the periodic interval in depth at which the horizontal cross section of the pit is measured. Each cross section is read in turn, consisting of 36 measurements of the radius from a central vertical reference line taken at 10° intervals in polar angle. For each radius r and x and z coordinates are calculated as follows

$$x_\theta = r_\theta \cos \theta \quad (x - \text{axis positive to the right on the screen})$$

$$z_\theta = -r_\theta \sin \theta \quad (z - \text{axis positive out of the screen})$$

and all y coordinates for the section are set to

$$y = -p \cdot (\text{depth interval})$$

where the y -axis is positive vertically up on the screen, and p is the number of the section ($p = 0$ for the surface section, 1 for the next section down, etc.).

The remaining points on the curved outline of the section are interpolated at 1° intervals from the 10° points. When all the cross sections have been specified, the base point coordinates are set up as $(0, -\text{max. depth}, 0)$. The display cycle is entered and the vertical section of the pit at 0° polar angle is displayed, with the full horizontal surface cross-section (although in this first display this appears as a horizontal line).

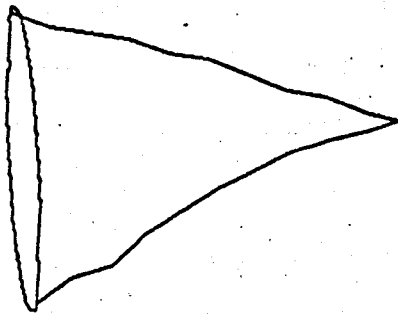
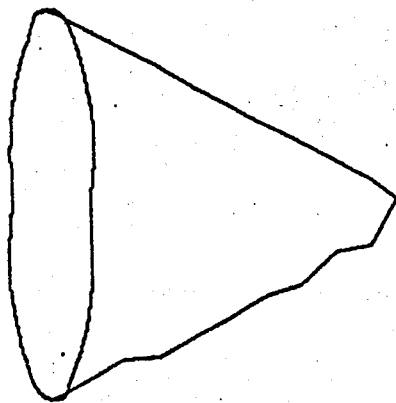
A series of light buttons are displayed allowing the three rotations XYROT, YZROT and ZXROT and the figure may be rotated as desired in 10^0 steps about any selected axis or axes by light pen "hit" on the light buttons. Procedures ROT and PERS are used in the display of the new figure. When the operator is satisfied with the display he may remove the light buttons and use FIPLLOT to copy the display file to the plotter. Alternatively by pressing function key 1 he may recall the light buttons, or by pressing function key 2 delete the whole display and read a new specification from paper tape. Light buttons are not continually written to the display file on each display cycle. The program uses procedures ROT and PERS.

A typical pit display is illustrated in Figure 3.22.

3.4.18

ARCJW027 Projectile point analysis (ALGOL 60 and EDGAR)

This program reads a closed curve specified by d-Mac coordinates with termination codes, using procedure READCLO - this is a modification of READFIG. The closed curve is the outline of a projectile point, and is preceded by a string giving its name, and the scale of the diagram. No vertical alignment points are specified as for READFIG. The name is printed and also punched on paper tape - since the system procedure OUTSTRING does not reproduce string quotes, it is necessary to produce them by CHAROUT if the paper tape produced is to be read by OUTSTRING later on. The curve is smoothed to eliminate hand-shake on the d-Mac pencil follower. The figure is then standardised by finding its areal centroid, then the point on the profile furthest away from this centroid (usually the point of the projectile). The whole figure is then rotated so that this furthest point is vertically above the centroid, and the distance from this point to the centroid is standardised, with adjustment of the scale factor.



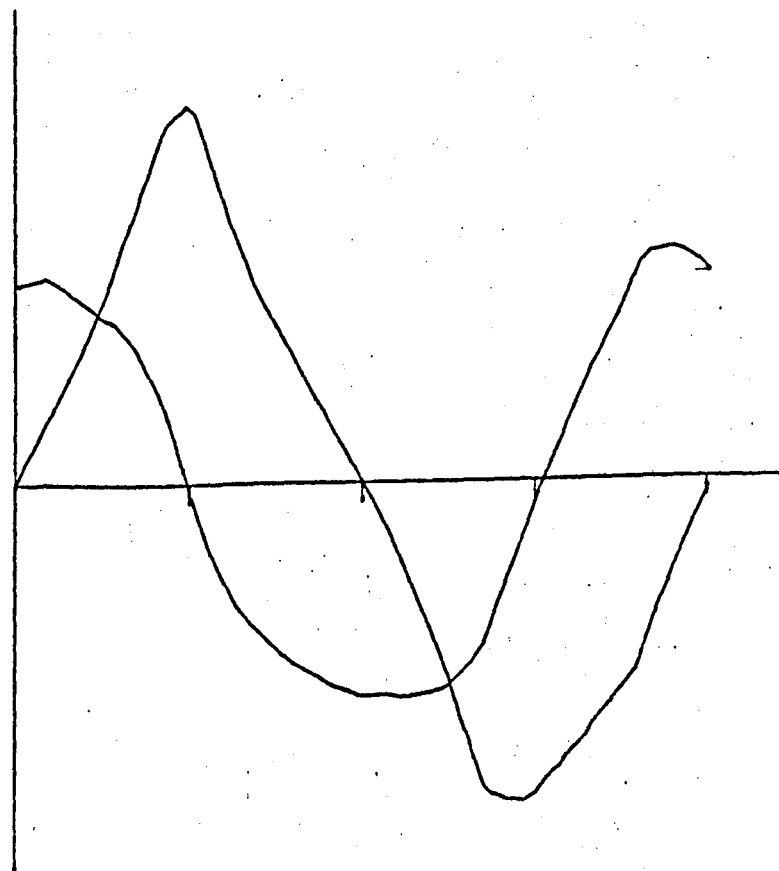
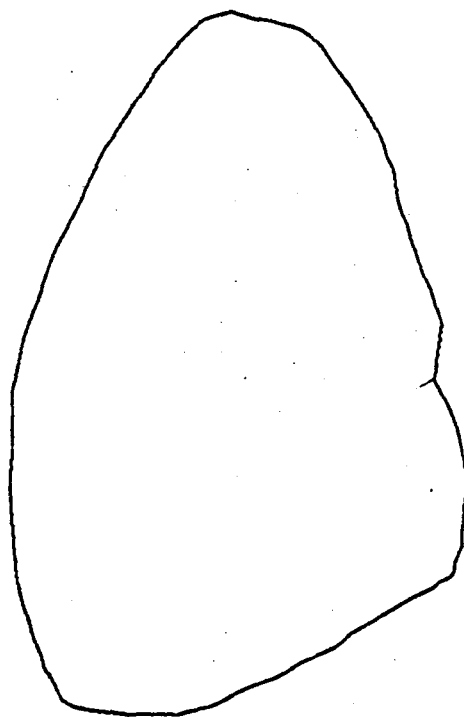
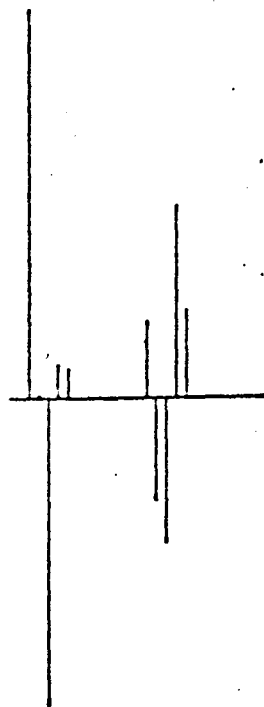
Two different displays of the same pit

Figure 3.22

The area is found as the summation of a number of elementary rectangular strips as the curve is followed from beginning to end, some of the strips being added to the summed area and some subtracted until the circuit is complete. The centroid is calculated during the same pass. During the rotation and standardisation pass each d-Mac coordinate has a polar angle allocated to it relative to the centroid, and there is a check that the polar angles always increase around the profile. These polar angles are used later in the Fourier analysis. Before exiting from READCLO the area and adjusted scale factor are output to both printer and paper tape.

The standardised closed figure is drawn on the display, identified by name and scale factor. The y and x traces of the figure are then plotted against polar angle, and it may be seen from Figure 3.23 that these approximate to sine and cosine curves respectively. Fourier analysis is then carried out on these curves up to order 5. The fourier coefficients for the y and x traces and their power spectra (sum of squared coefficients) are printed and also punched on paper tape. New y and x traces are synthesized from these coefficients and the resultant traces are displayed together with the synthesized outline of the closed figure. It is of course found that points are rounded and irregular outlines smoothed out considerably (see Figure 3.23). Perfect synthesis of the closed curve would require a complete Fourier transform. Experiments have been made with up to 10 orders; in some respects e.g. exaggeration of small irregularities in the profile the results were worse than the cruder profile generated by 5 orders, and in view of the success of similarity coefficients generated by 5 orders the extra calculation is not considered justified (see Figures 1.9 and 1.10).

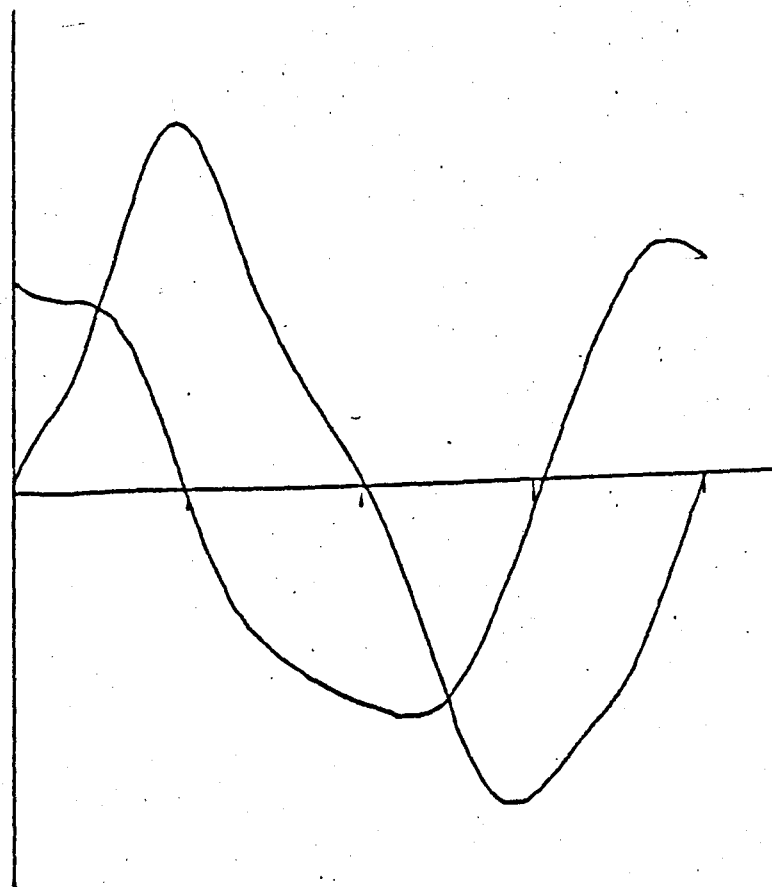
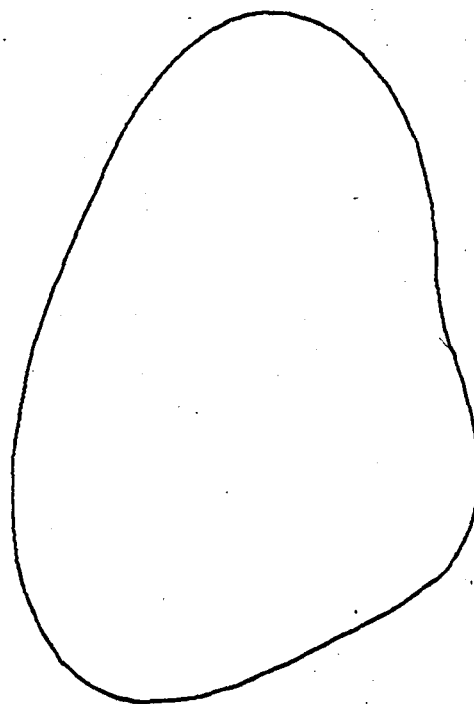
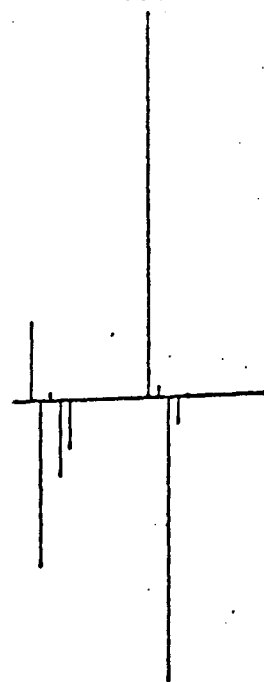
155



JEZERANY I 1

SCALE 1.37

125



Projectile point plot with Fourier analysis

Figure 3.23

The Fourier transform (first 5 orders) is plotted as a line spectrum on the screen, coefficients > 20 being truncated and labelled with their actual value (see Figure 3.23). A wait loop is then entered which requires a function key to be pressed - if this is done the display file is reset and another paper tape record is read, processed and displayed. A more usual course is, however, to call the plotter routine FIPLLOT to copy the display file in permanent form. ARCJW027 must then be called again to process a further record.

Termination characters may be punched on the paper tape by reading a dummy record consisting of a dummy name ('END') and scale, and one pseudo d-Mac coordinate X0000 Y0000. The punched paper tape is then in a form which may be read and processed by ARCJW028.

The program uses the following procedures:

CSIZE (plot the coefficient size on the screen; this is used for Fourier transform coefficients > 20), READCLO (read a closed figure, smooth it, find the areal centroid and standardise the distance from the centroid to the furthest point on the closed figure. Rotate the figure so that the standardised distance is vertical. The area is calculated as the summation of a number of rectangular strips. Calculate the polar angles and check that they always increase. Output the area and the adjusted scale factor).

This method of calculating profile parameters using Fourier analysis has not been seen elsewhere. The parameters may be used as a basis for comparing projectile points (see ARCJW028, paragraphs 1.10.6 and 3.4.19)

3.4.19

ARCJW028 Projectile point similarity (ALGOL 60)

This program uses the paper tape produced by ARCJW027 to calculate similarity coefficients between projectile points. The paper tape carries the Fourier coefficients and power series for five orders. The program is fully described in paragraph 1.10.6.

3.4.20

HISTOG Histogram plotter (BASIC - designed for use on remote access terminals)

This program plots a histogram on the teletype, as illustrated in Figures 1.12 and 1.13. It is not necessary to count the number of readings, nor to specify ranges and intervals. The program is fully described in paragraph 1.10.7.

3.4.21

PHENON Phenon ordering (BASIC - designed for use on remote access terminals)

This program orders a matrix of objects using similarity coefficients, producing the final order and phenon levels necessary for the construction of a dendrogram (e.g. Figures 1.18, 1.20 and 1.21). The program is fully described in paragraph 1.10.9.

3.4.22

SITPLT Site coordinates plot (BASIC - designed for use on remote access terminals)

When excavating a site it is necessary to record the position of every small find in 3D coordinates. This is done even when traditional methods are to be employed to draw site plans and sections, but is even more essential if the graphics terminal is to be used to produce 3D perspective diagrams, plans or sections of the site.

To obviate the need to take three measurements from datum lines on all occasions, a new method employing a theodolite has been developed for this study. The theodolite is set up on a tripod at

the site, centred over a reference point on the datum surface using a plumb line, levelled, and the height of the instrument above datum measured (d_i). The horizontal circle is rotated until the 0° mark is aligned with a major grid line of the site survey.

Pointing the telescope at the small find, the lead screws are adjusted until the cross-wires are centred on the find (it is useful to centre on the point of a white plastic marking-out skewer). The bearing (b_s) and depression (t_s) angles are recorded. A staff marked in m and cm is placed as near the find as possible, and the depth of the find below the instrument found by horizontal sighting (d_s).

Then the radius r of the small find from the reference point is given by

$$r = d_s / \tan t_s$$

and assuming that the 0° bearing is true north, bearing increasing clockwise, eastings and northings are given by

$$e_s = r \sin b_s = \frac{d_s \sin b_s}{\tan t_s}$$

$$n_s = r \cos b_s = \frac{d_s \cos b_s}{\tan t_s}$$

and the true depth $d = d_s - d_i$

The program allows angles to be read in degrees and minutes, taking, for example, the number 32.18 to mean 32 degrees 18 minutes. The fractional part of the number is corrected to a true fraction of a degree, e.g.

$$.18 \times \frac{5}{3} = .30 \text{ degree,}$$

added to the integer part of the number, then converted to radians.

Depression angles are checked to lie in the range

$$90.00 \leq t_s \leq 180.00$$

to accommodate the vertical circle readings, but these are converted to true depression angles by subtracting 90° . Bearing angles are checked to lie in the range

$$0.00 \leq b_s \leq 359.59$$

to accommodate the horizontal circle readings. The minute readings for both depression and bearing are checked to lie in the range

$$.00 \leq m \leq .59$$

The eastings, northings, depth below the instrument and a string identifier for each point are printed in a headed table. The program then returns to the reading phase, and a negative depth reading (i.e. specifying a point above the horizontal plane of the instrument) terminates the program.

To check the method a series of readings were taken with a staff and tapes accurately laid out on a hard courtyard surface. The calculations were made with both the above BASIC program and a similar ALGOL program, and found to be identical, hence no inaccuracy can be attributed to the trigonometrical functions in the BASIC compiler.

The experimental error was found not to exceed the following:

<i>Nominal Northings</i>	<i>Eastings error</i>		<i>Northings error</i>		
	<i>Absolute</i>	<i>Percentage</i>	<i>Absolute (from mean)</i>	<i>Error in mean</i>	<i>Percentage</i>
1.50	0.010	0.7	0.010	0.033	2.6
2.50	0.007	0.6	0.007	0.021	1.2
3.50	0.005	0.4	0.004	0.013	0.5

In general, if a distance $> 3m$ is maintained between the instrument and the small find, which will usually be the case, the error can be expected to be less than 1%, which is acceptable (1 cm error / m).

$$\begin{aligned}
 \text{Error in eastings} &= \frac{h + \delta h}{\tan(t - \delta t)} \cdot \sin(b + \delta b) - \frac{h \sin b}{\tan t} \\
 &= \frac{(h + \delta h)(\sin b \cos \delta b + \cos b \sin \delta b)(1 + \tan t \tan \delta t)}{(\tan t - \tan \delta t)} - \frac{h \sin b}{\tan t} \\
 &\approx \frac{(h \sin b \cos \delta b + h \cos b \sin \delta b + \delta h \sin b)(1 + \tan t \tan \delta t)}{\tan t - \tan \delta t} - \frac{h \sin b}{\tan t} \\
 &\approx \frac{(h \sin b \cos \delta b + h \sin b \tan t \tan \delta t + h \cos b \sin \delta b + \delta h \sin b)}{\tan t - \tan \delta t} - \frac{h \sin b}{\tan t}
 \end{aligned}$$

This expression gives the three approximate error terms

$$(h \sin b) \delta t, \left(\frac{h \cos b}{\tan t} \right) \delta b, \left(\frac{\sin b}{\tan t} \right) \delta h$$

To minimise the errors, it is necessary to keep t large, and h small i.e. mount the theodolite as close to the find as possible, and keep the theodolite as low as possible.

In the northings calculation, $\sin(b + \delta b)$ is replaced by $\cos(b - \delta b)$

$$\begin{aligned}
 \text{Error in northings} &= \frac{h + \delta h}{\tan(t - \delta t)} \cdot \cos(b - \delta b) - \frac{h \cos b}{\tan t} \\
 &= \frac{(h + \delta h)(\cos b \cos \delta b + \sin b \sin \delta b)(1 + \tan t \tan \delta t)}{\tan t - \tan \delta t} - \frac{h \cos b}{\tan t} \\
 &\approx \frac{(h \cos b \cos \delta b + h \sin b \sin \delta b + \delta h \cos b)(1 + \tan t \tan \delta t)}{\tan t - \tan \delta t} - \frac{h \cos b}{\tan t} \\
 &\approx \frac{(h \cos b \cos \delta b + h \cos b \tan t \tan \delta t + h \sin b \sin \delta b + \delta h \cos b)}{\tan t - \tan \delta t} - \frac{h \cos b}{\tan t}
 \end{aligned}$$

This expression gives the three approximate error terms

$$(h \cos b) \delta t, \left(\frac{h \sin b}{\tan t} \right) \delta b, \left(\frac{\cos b}{\tan t} \right) \delta h$$

To minimise the errors, it is necessary to keep t large, and h small as before. The value of b leads to errors in both eastings and northings. As b increases, two error terms in the eastings increase and one decreases, while one error term in the northings increases and two decrease. Thus errors in northings will predominate at small bearing angles, and errors in eastings will predominate at bearings near 90° . This is confirmed in practice by the above eastings & northings calculations which were at small bearing angles.

Typical runs of the program are shown in Figure 3.24.

EASTINGS	NORTHINGS	DEPTH	ITEM
.100731	1.52328	1.447	A1
.200122	1.52695	1.447	A2
.300502	1.52697	1.447	A3
.401629	1.52911	1.447	A4
.502075	1.53158	1.447	A5
.603122	1.53373	1.447	A6
.704019	1.53536	1.446	A7
.804462	1.53553	1.446	A8
.904944	1.53425	1.445	A9
1.00592	1.53624	1.445	A10
1.10689	1.53757	1.445	A11
1.20658	1.53697	1.445	A12
1.30689	1.5365	1.445	A13
1.40798	1.53835	1.444	A14
1.50994	1.53832	1.443	A15
1.60864	1.53457	1.441	A16
1.70808	1.53348	1.439	A17
.100365	2.51713	1.433	B1
.200518	2.51972	1.433	B2
.3001	2.51672	1.433	B3
.400389	2.51848	1.433	B4
.500516	2.5201	1.433	B5
.601496	2.52157	1.433	B6
.701189	2.51987	1.432	B7
.801339	2.51852	1.433	B8
.905321	2.52847	1.433	B9
1.00282	2.51999	1.433	B10
1.10359	2.52402	1.433	B11
1.20288	2.52	1.432	B12
1.30451	2.52385	1.432	B13
1.40574	2.5239	1.432	B14
1.50707	2.5248	1.432	B15
1.60643	2.52322	1.431	B16
1.70696	2.52276	1.429	B17
.100222	3.51474	1.423	C1
.200698	3.51633	1.423	C2
.300325	3.51516	1.423	C3
.401406	3.51401	1.423	C4
.501048	3.51323	1.423	C5
.601988	3.51562	1.423	C6
.701434	3.51566	1.423	C7
.80205	3.51604	1.423	C8
.900446	3.51125	1.423	C9
1.00108	3.51044	1.422	C10
1.10312	3.5129	1.422	C11
1.20307	3.51389	1.422	C12
1.30282	3.50943	1.422	C13
1.40294	3.51057	1.421	C14
1.50449	3.51308	1.42	C15
1.60448	3.5126	1.417	C16
1.70484	3.5136	1.415	C17
.653621	1.04196	1.23	CLAUDIAN COIN
.571485	.571486	1.52	SAMIAN DRAG. 33
1.09119	1.89001	1.26	SAMIAN DRAG. 27
.500001	.866029	1	TEGULA
1.00021	1.00022	2.45	IMBEX

READY

Graphics Bibliography

- Appel, A., 1968
Some techniques for shading machine rendering of solids
Proc. A.F.I.P.S. Spring Joint Computer Conference, 1968 32, 37-45
- Armit, A.P., 1969
A multipatch design system for Coons' patches
I.E.E. Conference Publication 51 152-161
International conference on computer aided design, Southampton
- Barney, G.C. and Hambury, J.N., 1971
The connection of storage display terminals to a timesharing computer
Computer Bulletin 15(1), 24-28
- Biddle, M. and Kjølbye-Biddle, B., 1969
Metres, areas and robbing
World Archaeology 1(2), 208-219
- Boehm, B.W., 1967
Tabular representations of multivariate functions - with applications to topographic modeling
Proc. A.C.M. National Meeting, 1967 403-415
- Burton, V., 1970
Computers confront the curator
Bulletin of the Metropolitan Museum of Art 1, 20-23
- Caulkins, D., 1972
GR1. Graphic Art
C.Hum. 6(3)
in *Humanities Programs Available* 180-181
- Comba, P.G., 1968
A procedure for detecting intersections of three-dimensional objects
Journal of the Association for Computing Machinery 15(3), 354-366
- Coons, S.A., 1964
Surfaces for computer-aided design of space figures
M.I.T. MAC-M-253 (Project MAC)
- Coons, S.A., 1967
Surfaces for computer-aided design of space forms
M.I.T. MAC-TR-41 (Project MAC)
- Denil, N.J., 1966
A language and model for computer-aided design
Proc. A.C.M. National Meeting, 1966 527-535
- Department of Trade and Industry
Computer Aided Design Centre, Cambridge, 1971
Computer art for the engineer
Computer Bulletin Supplement 15(7), vi

- d-Mac Limited *d-Mac Pencil Follower Type P.F. 10,000 Mark 1B
General Description and Specification*
d-Mac Limited, Glasgow 4 pp.
- Edwards, A.J., 1971 The industrial application of computer-
aided design
Computer Bulletin 15(1), 12-15
- Gerber Scientific Instrument Company, 1969
Gerber Graphics Generator Program (3G)
Prepared for the Gerber Scientific
Instrument Company, Hartford, Connecticut
by Applied Programming Technology Corporation
Sudbury, Massachusetts 27 September 1969
62 pp.
- Hall, D.J., Ball, G.H., Wolf, D.E. and Eusebio, J.W., 1969
PROMENADE - an interactive graphics pattern-
recognition system
Information Processing 68 951-956
North-Holland Publishing Company, Amsterdam
- Heap, B.R. and Laws, M., 1968 *Construction of the characters available
in the NPL graphical output system*
National Physical Laboratory Central
Computer Unit
Publication CCU 1, May 1968
- Hopgood, F.R.A., 1969 GROATS - A GRAPHIC Output package for ATLAS
using the SC 4020
The Science Research Council, Chilton,
Didcot, Berks.
- Hopgood, F.R.A., 1971 *Graphics section progress report*
The Science Research Council, Chilton,
Didcot, Berks.
- Johnson, T.E., 1963 Sketchpad III: a computer program for
drawing in three dimensions
*Proc. A.F.I.P.S. Spring Joint Computer
Conference, 1963* 347-353
- Jones, C.B., 1971 A new approach to the 'hidden line' problem
Computer Journal 14(3), 232-237
- Luzadder, W.J., 1968 *Basic Graphics*
Prentice-Hall, New Jersey 641 pp.
- MacCallum, K.J., 1970 Surfaces for interactive graphical design
Computer Journal 13(4), 352-358
- Nelson, P.M., 1967 *SC 4020 Plotter Manual*
The Science Research Council, Chilton,
Didcot, Berks.
- Notley, M.G., 1970 A graphical picture drawing language
Computer Bulletin 14(3), 68-74

- Opitz,H. and Simon,W., 1967 *EXAPT 1 and EXAPT 2 Part Programmer Reference Manuals*
Verein zur Förderung des Exapt-Programmier
systems e.V.
Aachen
- Ortony,A., 1971 A system for stereo viewing
Computer Journal 14(2), 140-144
- Patrick,E.A., Anderson,D.R. and Bechtel,F.K., 1968
Mapping multidimensional space to one
dimension for computer output display
I.E.E.E. Transactions on Computers
C-17(10), 949-953
- Pitteway,M.L.V., 1967 Algorithm for drawing ellipses or hypebolae
with a digital plotter
Computer Journal 10, 282-289
- Pitteway,M.L.V., 1972 The impact of computer graphics
Nature 235, 83-85
- Ricci,A., 1971 An algorithm for the removal of hidden lines
in 3D scenes
Computer Journal 14(4), 375-377
- Rosing,K.E., 1969 Computer graphics
Area 1, 2-7
Institute of British Geographers
- Schoenaker,G.C.M., 1968 Stereo drawings made with a digital computer
Philips Technical Review 29(3/4), 102-109
- Shepard,D., 1968 A two-dimensional interpolation function for
computer mapping of irregularly spaced data
Harvard papers in theoretical geography
"Geography and the properties of surfaces"
series 15 AD 668707 19 pp.
- Sutherland,I.E., 1963 Sketchpad: a man-machine graphical
communication system
Proc. A.F.I.P.S. Spring Joint Computer
Conference, 1963 329-346
- Taylor,F.E., 1971 Computer driven displays and the (BCS)
display group
Computer Bulletin 15(1), 4-11
- Treece,J.A., 1970 Realistic computer-generated displays
Electronics and Power 16 (June 1970),
205-207
- University of Arizona, 1971 Use of the University of Arizona CDC 6400
in anthropology
Newsletter of Computer Archaeology 6(4), 3-6

Chapter 4

POTTERY CLASSIFICATION

4. Pottery Classification

4.1

The computer has been used by some archaeologists to classify pottery. The usual method has been to record decoration, fabric, glaze and certain key measurements as attributes of a pot, then submit the data to factor analysis, matrix manipulation and other statistical methods (see Chapter 1). No attempt has been made to codify pottery profiles in an objective, systematic manner, however, and this thesis proposes a new method for the solution of this problem, employing the d-Mac pencil follower and treating the profile as multivariate data. The similarity coefficients are based on a form of Euclidian distance and are corrected for scale differences. An alternative measure, which enables "families" of pots of like profile to be detected, is not compensated for scale difference.

4.2

One of the earliest attempts to formulate a comprehensive geometric description of pottery vessels is given by A.O. Shepard (1957). The system proposed is an elaboration of the "aesthetic measure" of G.D. Birkhoff (*Aesthetic measure*, Cambridge, 1933) in which the points of a pot profile where the eye rests are determined. These points are *end points* (which determine the orifice diameter), *vertical tangents* (which determine the equatorial diameter), *points of inflection*, and *corner points* (i.e. carination or body/neck angle). The relative positions of these different types of point determine the form of the pot. Orifices may be classed as *unrestricted* (when the hands can be put in, the vessel is suitable for the display or drying of its contents, and the orifice diameter is the maximum diameter of the vessel), *restricted* (when the vessel is suitable for storage

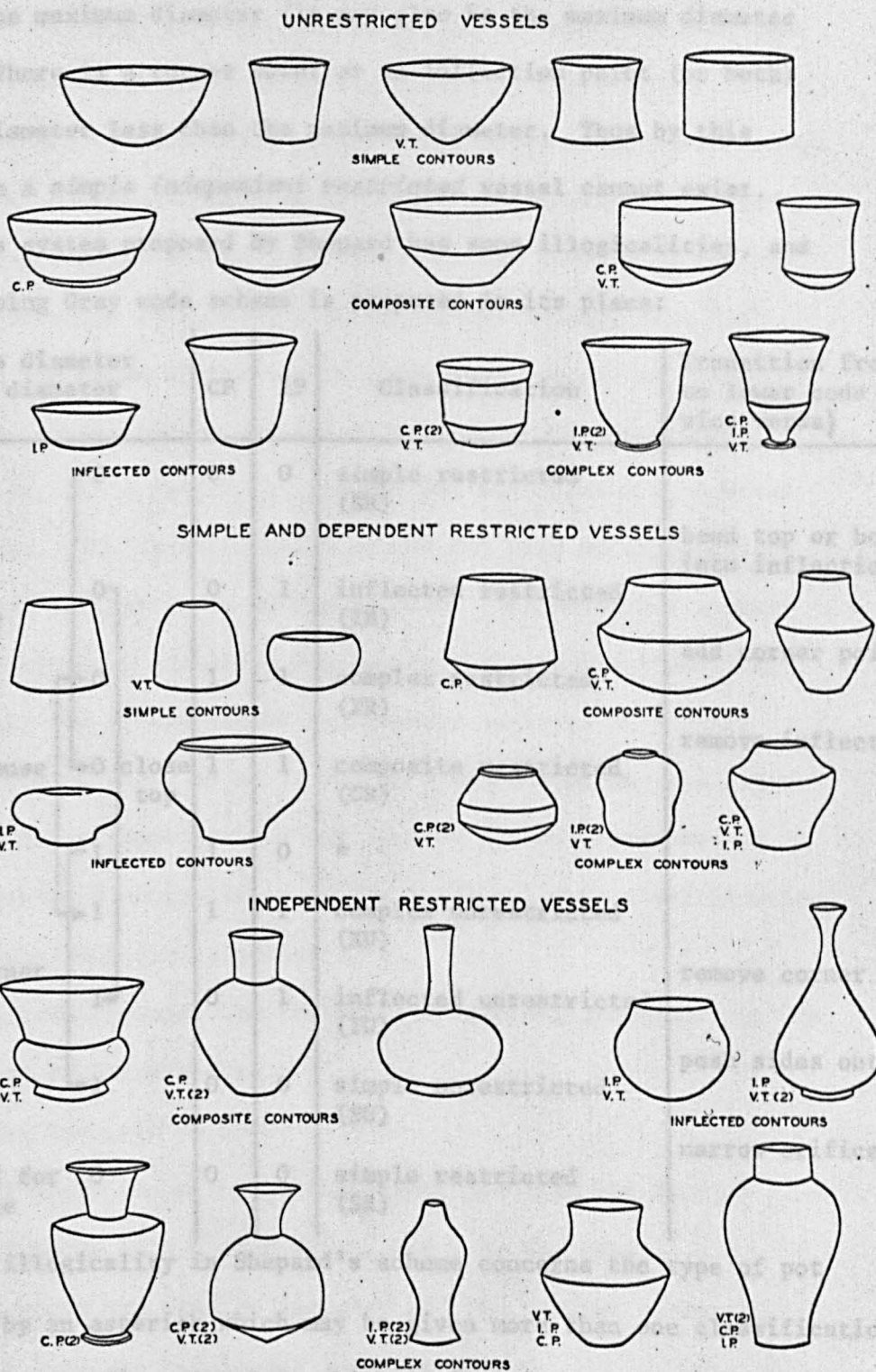
and protection of its contents and the orifice diameter is less than the maximum diameter), or *necks* (preventing slopping, and facilitating pouring of the contents). For a restricted vessel, as the position of the vertical tangent approaches the orifice the vessel becomes less and less restricted, and the transition between restricted and unrestricted form occurs when the vertical tangent is at the orifice. The diagram illustrating Shepard's types is reproduced in Figure 4.1.

Unrestricted vessels may further be classified as *simple*, *composite*, *inflected* or *complex*. A composite vessel has a corner point but no inflection point. An inflected vessel has a point of inflection but no corner point, and a complex vessel has both corner and inflection points. This may be expressed by the Boolean table:

Classification	CP	IP
simple	0	0
inflected	0	1
composite	1	0
complex	1	1

The vessel always gets wider from base to orifice.

Restricted vessels may be classified similarly, but here another factor must be considered, that of dependency. *Dependent restricted* vessels have an orifice diameter less than the maximum diameter, and have no restriction (of less diameter than the orifice) marked by a corner point or inflection point. Any corner points which exist are at larger diameters than the orifice. The dependency shows itself by the fact that the profile must curve towards the orifice and base from the maximum diameter. *Independent restricted* vessels have restrictions either at the orifice or some



Shepard's classification system

Figure 4.1

other point, but the profile may convolute with no dependency on other parts of the profile. The orifice diameter is not dependent on the maximum diameter (it may also be the maximum diameter itself). There is a corner point or an inflection point (or both) at some diameter less than the maximum diameter. Thus by this definition a *simple independent restricted* vessel cannot exist.

This system proposed by Shepard has some illogicalities, and the following Gray code scheme is proposed in its place:

Orifice diameter = max. diameter		CP	IP	Classification	Transition from upper to lower code (may be vice versa)
add corner point	0	0	0	simple restricted (SR)	bend top or bottom into inflection
	0	0	1	inflected restricted (IR)	
	0	1	1	complex restricted (XR)	add corner point
open top, reverse sense of CP if necessary	0	1	1	composite restricted, (CR)	remove inflection
	1	1	0	*	
remove corner point	1	1	1	complex unrestricted (XU)	remove corner point
	1	0	1	inflected unrestricted (IU)	
	1	0	0	simple unrestricted (SU)	push sides out
duplicated for 0 convenience	0	0	0	simple restricted (SR)	narrow orifice
	0	0	0	simple restricted (SR)	

The illogicality in Shepard's scheme concerns the type of pot indicated by an asterisk which may be given more than one classification, depending on the direction from which it is approached. From the CR type the neck may be expanded until it becomes the maximum diameter,

and this would then be called an independent composite restricted vessel by Shepard. From the XU type the inflection may be removed, producing a composite unrestricted vessel in Shepard's terminology. Both these types have their maximum diameter at the orifice, have a corner point and no inflection point. Thus there is a discontinuity in what is a continuous spectrum of shapes with smooth transitions. All the above transitions may take place in the reverse direction. One extra operation is necessary at the discontinuity, viz. the reversal of the sense of the corner point from concave to convex or vice. versa. Unrestricted vessels must necessarily have convex corner points, while restricted vessels may have convex or concave corner points. The classification scheme may also be represented on a *Karnaugh Map*, as shown in Figure 4.2a. The adjacency properties of the minterms admirably represents the possible transitions. Alternatively the scheme may be represented as nodes of a cubic graph (see Figure 4.2b) or as states in a *state diagram* (see Figure 4.2c). The transitions between the various states are caused by different kinds of manipulation of the profile. This classification scheme has not been seen elsewhere.

The scheme suggests the following diagnostic procedure for the determination of pottery type:

- (a) set the type to zero;
- (b) if there is an inflection point, add 1 to the type;
- (c) if there is a corner point, add 2 to the type;
- (d) measure the orifice diameter and the maximum diameter;
if they are equal, add 4 to the type.

CI	M		
		0	1
	00	SR	SU
	01	IR	IU
	11	XR	XR/XU
	10	CR	CR/CU

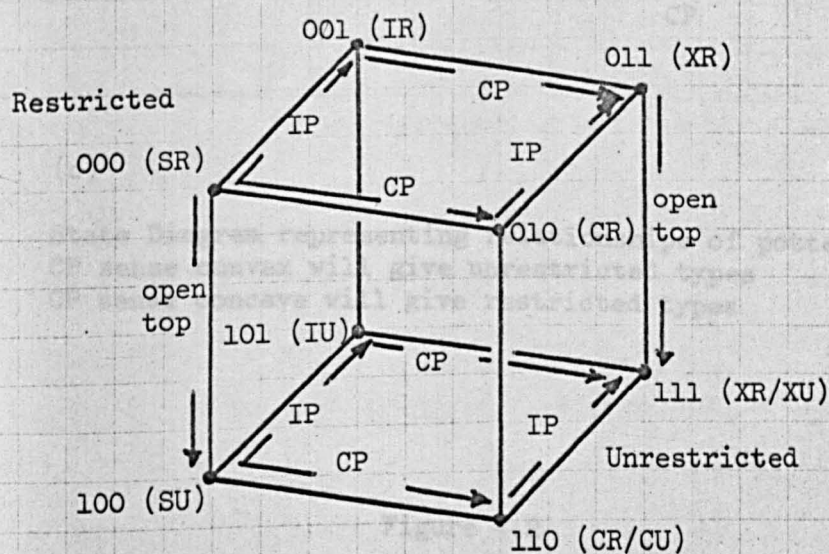
(a)

Karnaugh map representing adjacencies for pottery types (the map is cyclic, square 000 SR is adjacent to square 010 CR and square 100 SU is adjacent to square 110 CR/CU).

M = orifice diameter is maximum diameter

C = corner point present

I = inflection point present



(b)

Cubic graph representing adjacencies for pottery types.

Figure 4.2

Present State	Next state			Alter CP sense
	Alter binary sense of:			
	M	C	I	
SR	SU	CR	IR	-
IR	IU	XR	SR	-
XR	XR/XU	IR	CR	-
CR	CR/CU	SR	XR	-
CR/CU	CR	SU	XR/XU	CU
XR/XU	XR	IU	CR/CU	XU
IU	IR	XR/XU	SU	-
SU	SR	CR/CU	IU	-
CU	CR	SU	XU	CR/CU
XU	XR	IU	CU	XR/XU

(c) continued

State Table representing the relationships between pottery types

Figure 4.2

Sheet 3 of 3 sheets

The type number expressed in binary then represents the pottery classification from the above table.

4.3

Other work on pottery classification which is relevant to the present study is given below.

4.3.1

Freeman (1961) discusses the encoding of arbitrary geometric configurations, and the transmission of geometric information over a telephone line. Although Freeman does not mention pottery profiles, most of his techniques are applicable to them.

One method is to cut the curve into arbitrary segments and approximate each by a standard mathematical curve. If a small number of segments are used, high-order polynomials become necessary, while short straight lines suffice if a large number of segments are used. The second approach has been used in this thesis, since plotters and display units are well suited to the drawing of straight lines, and the large number of coordinates generated by the d-Mac pencil follower readily allow the curve to be specified as a large number of segments. Freeman suggests that a non-uniform length of segment may give more efficient coding. This technique has also been employed in this thesis by taking a large number of coordinate points around a curve of large curvature, while reducing the number for curves of small curvature on the d-Mac diagram.

Freeman's paper is limited to the use of straight-line segments, using either a single length or a small number of standard lengths. He discusses the use of *parametric curves*, for which the variations of the x and y coordinates are expressed as functions of the length of curve traversed. The two parametric curves are then submitted to Fourier analysis and the frequency spectra are examined. In general, the

spectra will be infinite, with progressively-increasing attenuation of the higher-frequency components. The *sampling interval* which will give adequate representation of the curve is less than half the wavelength of the highest significant frequency component. Once the sampling interval has been determined, the curve can be described as a sequence of angles at which successive segments are placed. Standard angles (say at 18° intervals, allowing 20 different orientations in the 360° circle) may be used, and by making the sampling interval small precision may be obtained (the "*unit vector method*"). Fourier analysis has been used in this thesis to describe projectile points, which are examples of closed curves (see paragraphs 3.4.18, 3.4.19 and 1.10.6), but the treatment is quite different to Freeman's method.

A second method described by Freeman uses a rectangular grid, and the curve is specified by the grid points which are nearest to the curve. This is the method used by digital incremental plotters using the 8-vector format. A hexagonal grid may also be used. If standard angles are used in the specification of curves, binary codes may be used to indicate the addition or subtraction of a standard angle between the points of a curve. Generating binary functions are given for a number of repetitive sequences. Freeman notes that contraction of the scale of a curve must introduce some distortion, since the binary code method then becomes less precise. An algorithm is given for finding the area of a closed curve specified in this system. Although Freeman only considers 2D curves, the techniques could be extended to 3D.

4.3.2

Clarke (1962, 1970) describes British Beaker pottery, He expresses pottery shape as ratios of various measurements to the waist

diameter. Five such ratios are specified, each subdivided into ranges (the choice of which is influenced by the number of vessels in each class), yielding 16 variables, and another 7 variables describe rim shape, curvature, etc., making 23 variables in all. Decorative motifs, the position of the decoration, and the paste and firing were also considered as a basis for classification, although the last was found to be remarkably consistent and was therefore omitted. 39 variables in all were cross-correlated in a 39 x 39 attribute - attribute square matrix and submitted to matrix analysis using the EDSAC II at Cambridge. The matrix is sorted so that high values of correlation appear near the central diagonal. Peaks or groupings of attributes are evident in the sorted matrix, and Clarke has taken these to define various groups of Beaker pottery types. The results are in general sensible, and several "intrusive" types are defined, although the method has been severely criticised by Matthews (see paragraph 1.4.8). The pottery profile analysis carried out in this thesis (see paragraph 4.4.3) shows that most of these intrusive types may be isolated on shape alone, although there are discrepancies which indicate that Clarke has attached more weight in some of his classifications to decoration and aspects other than shape (see paragraph 1.10.1 and Figures 1.3, 1.4).

In his 1970 work, Clarke studies and classifies all known complete examples of British Beaker pottery, and his new system of classification mentioned above is proposed to supersede that of Thurnam and Abercromby. Volume I outlines the basis of classification and discusses the analysis of the material. The second volume contains line drawings of all known complete examples, with detailed inventories. Chapter 3 of Volume I is largely a repetition of the

1962 paper. An appendix gives histograms for rim/waist, belly/waist and foot/waist diameter ratios, and for rim height/waist height and rim height/waist diameter ratios. Mutually-exclusive traits are noted by a special symbol to distinguish these pairs from pairs of traits which could occur together but never do so.

4.3.3

Sorenson (1964) describes the use of automated tools in archaeology. By a reversal of the normal NC machine tool (the APT programming language for such machines has been described in paragraph 3.3.5), a stylus guided over the surface of an object may be used to automatically compute the dimensions of the object and record them on tape. The tape would then normally be used to manufacture precise duplicates of the object, but alternatively may be read by computer and used for the description of the object. A second device shapes plastic relief maps when the operator traces a stylus over contours on a 2D map. If a photocell were to replace the stylus perhaps photographs could also be used for the automatic measurement of objects. This was first suggested by G. Carlson of the Computer Research Center, Brigham Young University, Provo, Utah, U.S.A. To counter the suggestion that this is "a dangerous intrusion of the dehumanizing machine into the domain of scholarship", Sorenson maintains that machines should be used for whatever they can do better than man alone, and that machines can free the archaeologist from the tedium of measuring sherds, so that his higher conceptual skills can be put to use for a greater proportion of the time.

4.3.4

McGimsey and Green (1965) of the University of Arkansas Museum present a coding scheme for ceramics to be used on IBM computers.

4.3.5

McPherron (1966) reports FORTRAN software for the IBM 7090 machine to study the variability in pottery style. The program uses attribute clustering, and attempts to define sociocultural differences based on the variability of samples.

4.3.6

In a series of papers (Rottländer, 1966, 1967, 1969; Holzhausen and Rottländer, 1970) Rottländer investigates provincial Roman pottery, deducing that the potters used measures of various sorts. Statistical techniques are used to find the most likely units of measurement employed to standardize the pots. The decorative collar on form Drag.38 is deduced to be carefully designed to allow the stacking of these expensive vessels during transport. The Roman potter evidently knew the extent of shrinkage during firing, and was able to control it to the convenient value of $\frac{1}{8}$. It seems that the ancient measuring units were of Greek origin and were adopted by Roman craftsmen. The same feature applies in the case of mosaic design; (see Moore paragraph 6.8.2).

4.3.7

Rutovitz (1966) gives a good survey of pattern recognition methods applied to character recognition. Edge-tracing methods are often used, and the problems are similar to those encountered in pottery profile recognition. Small bumps on boundaries often give trouble. This trouble has been partly removed in the pottery profile study in this thesis by smoothing the d-Mac profile by computation (see paragraph 4.4.3). Curvature is discussed, together with the detection of a "corner", viz. a succession of points for which the local curvature rises above a certain threshold.

A *shrink algorithm* is defined using a "*crossing number*" which gives the number of changes in binary value as the eight neighbour elements of a given element are scanned in sequence (1 = black, 0 = white).

An element is changed to 0 if

(a) it is originally 1;

AND (b) the crossing number of the element is 2;

AND (c) there are at least two black neighbours;

AND (d) at least one of the two horizontal neighbours or the upper vertical neighbour is white,

OR the crossing number of the upper vertical neighbour is not 2;

AND (e) at least one of the two vertical neighbours or the right-hand horizontal neighbour is white,

OR the crossing number of the right-hand horizontal neighbour is not 2.

Condition (b) prevents the removal of thin lines and hole formation. Condition (c) preserves the end points of thin lines. Conditions (d) and (e) prevent the separation of connected subsets of the picture. This shrink algorithm could be used in the numerical definition of a thin line profile from a binary matrix representation of a conventional pottery section, formed by automatic *flying spot scanner*, the *Nipkow Disc* or other scanning methods used for character recognition in computers.

Rutovitz further discusses hierarchical systems, Backus notation and recursive notation in pattern recognition, and the description of pictures in simple languages, e.g. in heraldry.

4.3.8

Ankel (1966, 1967) reports software for the analysis of prehistoric pottery. The system is named ARDOC (see Gundlach,

paragraph 2.3.7) and is an information retrieval scheme. A 10mm grid is used to record Bronze Age vessels in three planes. In each plane, the positions where the profile crosses the grid lines are noted, and it has been found possible to store and reproduce on command visual representations approximating to the actual shapes of the pottery. A mm grid has been used in this thesis for a different application, the calculation of volumes (see paragraph 4.4.5).

4.3.9

Gardin (1967) gives methods for the descriptive analysis of archaeological material. He gives profile codes for pottery and other artefacts. Gardin does not assume that scientific standards have been used in the formulation of the data upon which algorithms are to be performed for information retrieval and classification purposes. Instead he gives examples of descriptive codes used by himself since 1955 for the storage of archaeological data on punched cards. To overcome the shortcomings of natural language (without scientific basis) Gardin states that three categories of rules are required, concerning orientation, segmentation and differentiation. His concluding remarks concern the relation of the descriptive languages so obtained to scientific language in general. It is now generally realised that the computer must be used to analyse large quantities of archaeological data, with the result that far more objective and scientifically-based measurements must be taken than those described by Gardin.

4.3.10

Sèvres porcelain has been analysed by Dauterman (1968) in the U.S.A. and by Stones and Williamson (1970) in Britain (the latter in consultation with Wilcock). The Sèvres porcelain factory of the 18th

and 19th centuries kept very careful records of payrolls and sales. In both projects the computer has been used to correlate various combinations of diacritical markings, colours, finishes, purpose and other attributes of particular porcelain objects. It is possible by this means to make discoveries concerning the dating of the pieces, the names of the artisans and the names and addresses of the purchasers.

4.3.11

Freeman (1968) describes computer methods for the processing, classifying and matching of profiles and other irregular curves, a technique which potentially enables whole vessels to be constructed from a pile of potsherds. The example given concerns the assembly of cut-out pieces of a complex crossword puzzle, but the method is applicable to the fitting together of irregular surfaces of any kind.

4.3.12

Rowlett (1968) describes the work of Berard (1883-1918), which included a classificatory system for early La Tène Marnian ceramics (carinated and tri-conic forms).

4.3.13

The I.E.E. Conference on pattern recognition (1968) contains a number of papers of possible application to the recognition of pottery profiles, including a paper by Aleksander and Albrow concerning the use of adaptive logic circuits.

Watanabi (1969) discusses the use of clustering in pattern recognition. The clustering is indicated on a scalogram using "rubber band" loops with circular arcs at the extremities. There are references to the method of Shepard and Kruskal. A *monte carlo* method for the recognition of a closed curve is described. Straight lines of random orientation and position are generated and the lengths of the

intersections with the closed curve are averaged. The average intersection length is proportional to the enclosed area, and thus the method is useful only in discriminating between figures of different areas. However, the method is independent of the orientation of the figures. This method could be applied to projectile points (see paragraphs 3.4.18 and 1.10.6) but would only discriminate on the basis of area, which is an insufficient criterion.

Rosenfeld surveys picture processing by computer in two 1969 publications. The treatment is 2D only. The material in the two publications covers the same topics, but the Academic Press book is the more theoretical in style. Picture compression, redundancy and efficient encoding is covered. Line drawings can be encoded very compactly, particularly if they involve sections consisting of straight lines or conic sections. Lines may also be chain encoded, using the nearest grid points. If the number of points has to be limited, it is better to put most points at the high-curvature parts of the curve. A matrix may also be used for indicating connections between the nodes of a line network. Each element e_{ij} indicates the number of arcs between node i and node j . The idea of template matching in the spatial frequency domain, i.e. by making comparisons of Fourier transforms rather than the actual pictures, is put forward. An idea similar to this has been employed in this thesis for the comparison of projectile points (see paragraphs 3.4.18 and 1.10.6). Filtering may also be employed to enhance pictures; this uses an inverse Fourier transform and, for example, may remove low frequencies (high-pass filter) or high frequencies (low-pass filter). High-pass filtering preserves sharp edges and fine detail, while washing out smooth regions. Low-pass filtering blurs the edges. By considering only low frequencies (as has been done in the projectile point analysis) a low-pass filter has been employed. Edges can be

expected blurred and sharp points removed. One method of comparing two features is to multiply the two Fourier transforms then take the inverse Fourier transform. In the projectile point analysis individual coefficients have been compared and weighted according to the inverse of the power series (see paragraph 1.10.6). Filtering is also of application in survey reduction for the removal of noise. This is covered in chapter 5.

4.3.14

Gaines (1970) discusses the description and analysis of ceramics by computer. She finds that the general problems encountered are the lack of standardisation of terms and concepts, insufficient records of attribute characteristics, the use of ceramic "types" and the definition of terms. Gaines favours the use of mnemonic codes input on cards, being of the opinion that it is better to use coded information than plain language to save time. This is a valid viewpoint, but the archaeologist without computer training will always prefer to see plain language output. Plain language has been employed wherever possible in the information retrieval applications of this thesis (see paragraph 2.6.2), but it is interesting to note that when archaeologists become involved in the preparation of data they soon decide to use abbreviations. This has indeed happened in the Doncaster project (see paragraphs 2.6.12 - 2.6.17). The plain language check files illustrated in Figures 2.9 and 2.10 have recently been revised by P. Buckland to include abbreviations and mnemonics.

Gaines describes the analysis of known pottery types to produce statistical profiles; unknown sherds may then be matched with these profiles. The software is written in FORTRAN for the GE 225 machine. The χ^2 test is used to establish the significance

of the attribute profiles. The correlation of the computer-derived types and traditional types is found to be good.

4.3.15

Orton (1970, 1971) performs statistical analyses on a collection of pottery from a Romano-British kiln site in Highgate Wood. Mr Orton was one of the prize-winners in a recent New Scientist - GEIS competition and was awarded the use of a remote terminal for a year. The terminal has been used for the progressive analysis of the excavation data. Cumulative percentage graphs (the shortcomings of which have been described in section 1.7) have been used to analyse pottery types. The Kolmogorov-Smirnov test (see paragraph 1.7.1) is used on these distributions. For two cumulative percentage graphs, the difference maximum D is significant for sample sizes n and m (both over 30) if it is greater than

$$1.36 \left((n+m)/nm \right)^{\frac{1}{2}} \text{ at the 95\% confidence level}$$

$$\text{or } 1.63 \left((n+m)/nm \right)^{\frac{1}{2}} \text{ at the 99\% confidence level.}$$

For $n = m = 100$ these expressions give critical values for D of 0.19 and 0.24 respectively ($100\% = 1.0$).

If Y is a matrix describing the number of pottery bases of each type in each lot of pottery and a matrix X similarly describes the pottery rims, then $Y = XP$ where P is a probability matrix relating the rims and bases; a proportion p of pots with rim x have base y . Multiplying both sides by X' we have

$$X' Y = X' X P$$

$$\therefore P = \frac{X' Y}{X' X}$$

This equation is formally identical with that for estimating regression coefficients in multiple linear regression, and suggests that this may be a pertinent technique. The reconstruction of the

pottery forms is to be attempted using this method.

4.3.16

Jardine (1971) describes a clustering approach to pattern recognition (cf. Watanabi, 1969, paragraph 4.3.13), in which the classes depend on the population, the method which has also been developed for this thesis since 1967. The cluster approach was first suggested by Sebestyen, Ball and Duda in 1962. Jardine gives examples of the recognition of hand-written characters by clustering; it is evident from the argument that he assumes migration is possible, although this is not stated explicitly. Examples of pottery clusters may be seen in Figures 1.1 - 1.4 and 1.18 - 1.20.

4.3.17

Perlman and Asaro (1971) describe pottery analysis by neutron activation. Powdered pottery is compacted into a pill with cellulose, then radiated in an atomic pile. The resultant activations are analysed, and the results submitted to statistical calculations by computer. A trial group is first compiled. Every member of the group is then tested for agreement with the group mean. Those items no longer in sufficient agreement with the group are discarded, and new group means are computed. Finally a representative group is obtained.

4.3.18

Irwin, Hurd and La Jeunesse (1971) describe a computer system for the input of artefact data without normal data preparation methods and conventional measurements on the artefacts. The equipment consists of an IBM 360/67 main frame with an Interdata Model 15 as "front-end" computer controlling a "Graf/Pen" graphics tablet, magnetic tape unit and teletype. All digital information about the artefacts are input through the tablet (for examples see Figure 3.3).

Scales are specified by touching a programmed interval with the stylus, such as 10cm, and all measurements are then read in hundredths of this unit. Using a plan aerial photograph of a site, accurate measurements of structures can be made simply by touching in a different scale. The technique is applicable to all archaeological objects, particularly to pottery profiles (from photographs). It is to be used in the study of skulls to investigate the relationship between *Homo neanderthalensis* and *Homo sapiens*. It is stated that 1000 measurements take about 1 hour using the technique (about one measurement every 3.6 seconds).

4.3.19

Wagner (1971) describes recent work at Tell-Hesi in Palestine. Every recorded piece of pottery is allocated a coded description, but an optional plain-language printout of the coded version is available. The computer is used to perform ranking, cluster analysis and principal component analysis on the data.

4.4

Pottery Classification Programs

Pottery classification programs have been written for the calculation of volumes of wheel-made pots, the calculation of profile parameters and the production of pottery diagrams for direct publication. Full listings of the programs are available in the printout file.

4.4.1

ARCJW005 Pottery Volumes (ALGOL 60)

This program is based on a process evolved for the estimation of pottery volumes using a mm grid and a hand calculator. The left-hand inner profile of the pot (at any scale) is superimposed on the grid. If the height of the pot is greater than or equal to the maximum radius the pot is designated as "tall type" and the estimation of volume is as a series of cylindrical shells of thickness 2mm. Thus the small dimension is stepped in constant increments while there is less error in the measurements of the heights which are of the order of cm. Readings of the cylinder height are taken from the top of the pot (or upper intersection of the profile) to the lower intersection of the profile at positions 1mm, 3mm (2n-1)mm from the centre vertical. The heights in mm are multiplied by their corresponding radii in mm, summed, and finally multiplied by $\frac{2 \times 2\pi \times S^3}{10^6}$ to give the volume in

litres, where S is the scale multiplier, i.e.

$$V = \frac{4\pi S^3}{10^6} \sum_{i=1}^n (2i-1)h_i$$

If the height is less than the maximum radius the pot is designated as "dish type" and the estimation of volume is as a series of flat discs of thickness 2mm. Thus again the smaller dimension is stepped in constant increments starting 1mm from the top of the pot while there is less error in the measurements of the radii which are of the order of cm.

The squares of the radii in mm of the elementary discs are summed then multiplied by $\frac{2 \times \pi \times S^3}{10^6}$ to give the volume in litres

i.e.
$$V = \frac{2\pi S^3}{10^6} \sum_{i=1}^n r_i^2$$

In the computer implementation, the inner profiles of the pot are recorded as coordinates using the d-Mac pencil follower. The profile is smoothed to remove hand-shake and rotated to a truly vertical position, then scaled to actual size. The profile curve is then followed to determine the heights of the elementary cylindrical shells or radii of the elementary discs as applicable, then the summation and correction is performed. The input is the d-Mac paper tape and there is line printer output. Figures 4.3 and 4.4 illustrate typical hand calculator printouts for tall type and dish type pots respectively. Figure 4.5 shows typical pot profiles showing tall and dish type estimations. The error in this method is found to be not greater than 5% if measurements are taken to the nearest mm. The d-Mac pencil follower takes measurements to 0.1mm and the accuracy can be expected to be better than 5%. The reason for the two methods is that the larger the measurements that are taken, the less will be the percentage error. Thus if the larger dimension (height or radius) is measured and the smaller stepped at a fixed interval, the error may be kept low.

4.4.2

ARCJW018 MDSCAL Disc (FORTRAN)

This multidimensional scaling program has been used to produce scalograms of common Samian types and British Beaker pottery (see Figures 1.1 - 1.4). The data for these runs consists of half matrices less diagonal of similarity coefficients for each pair of pots, calculated using the method described in paragraph 4.4.3 below.

1 x <	Elementary cylindrical
4 6 =	shell of radius 1mm
4 6 S	and height 46mm
3 x <	Elementary cylindrical
4 7 =	shell of radius 3mm
1 8 7 S	and height 47mm
5 x <	
4 4 =	
4 0 7 S	
7 x <	
4 3 =	
7 0 8 S	
9 x <	
4 2 =	
1 0 8 6 S	
1 1 x <	
3 1 =	
1 4 2 7 S	$\frac{1}{4}$ scale pot
1 3 x <	(Scale factor s = 4)
2 4 =	
1 7 3 9 S	
1 5 x <	
1 9 =	
2 0 2 4 S	
1 7 x <	
1 5 =	
2 2 7 9 S	
1 9 x <	Last elementary cylindrical
9 =	shell of radius 19mm
2 4 5 0 x T	and height 9mm
8 0 3 8 4 =	Correction factor = $4\pi s^3/10^6$
1.9 6 9 4 0 8 0 0 S	Volume in litres

Figure 4.3

Hand calculator printout for "tall type" pot
volume calculation

37 x < Elementary disc
37 = of radius 37mm
1369 S

37 x <
37 =
2738 S

36 x <
36 =
4034 S

35 x <
35 =
5259 S

33 x <
33 =
6348 S

31 x <
31 =
7309 S

1 scale pot

30 x < (Scale factor s = 4)
30 =
8209 S

28 x <
28 =
8993 S

25 x <
25 =
9618 S

22 x <
22 =
10102 S

18 x <
18 =
10426 S

10 x < Elementary disc
10 = of radius 10mm
10526 S

37 x <
37 =
11895 S

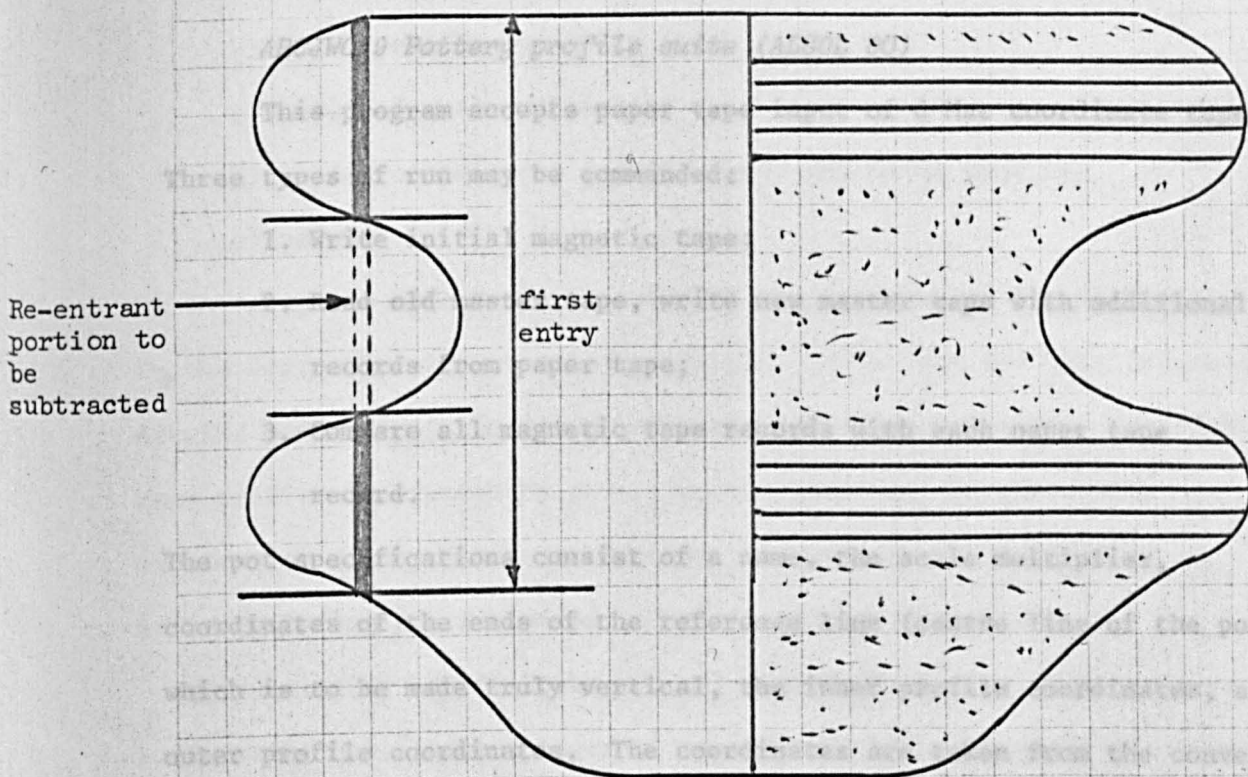
11895 T

11895 x <
40192 =

4.78083840 S Correction factor = $2\pi s^3/10^6$
Volume in litres

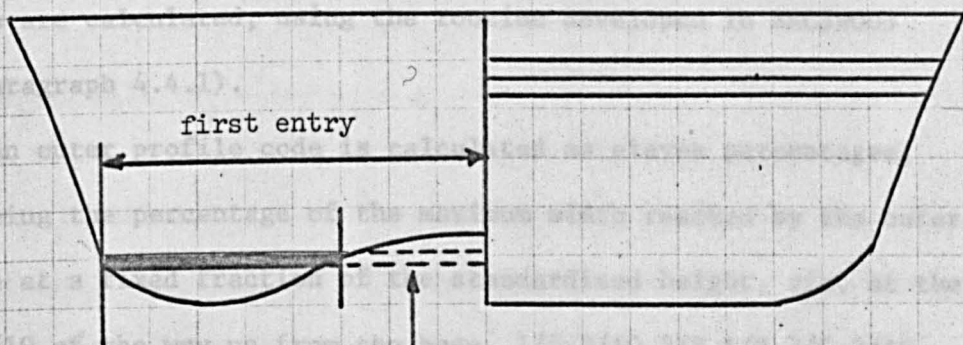
Figure 4.4

Hand calculator printout for "dish type" pot volume calculation



"Tall type" pot

Volume estimated as a series of cylindrical shells



Inner disc to be subtracted

"Dish type" pot

Volume estimated as a series of discs

Figure 4.5

Estimation of pottery volumes

The program is fully described in paragraph 1.10.1.

4.4.3

ARCJW019 Pottery profile suite (ALGOL 60)

This program accepts paper tape input of d-Mac coordinate tapes.

Three types of run may be commanded:

1. Write initial magnetic tape;
2. Read old master tape, write new master tape with additional records from paper tape;
3. Compare all magnetic tape records with each paper tape record.

The pot specifications consist of a name, the scale multiplier, coordinates of the ends of the reference line (centre line of the pot) which is to be made truly vertical, the inner profile coordinates, and outer profile coordinates. The coordinates are taken from the conventional archaeological pot drawing with left-hand section, right-hand elevation. Beginnings and ends of the profiles are denoted by dummy coordinates X0000 Y0000 and X0001 Y0001 respectively, a subset of the code developed for general straight-line figures (e.g. Masons' marks, see paragraph 1.10.3). The coordinates are rotated, smoothed and scaled to a standard height (with adjustment of the scale factor). Volumes are calculated, using the routine developed in ARCJW005 (see paragraph 4.4.1).

An outer profile code is calculated as eleven percentages, each being the percentage of the maximum width reached by the outer profile at a fixed fraction of the standardised height, viz. at the base, 1/10 of the way up from the base, 1/5, 3/10, 2/5, 1/2, 3/5, 7/10, 4/5, 9/10 way up from the base, and at the top. This is an elaboration of the height/maximum width ratios used by other pottery classifiers.

The following items are written to magnetic tape for each

pot: a) name

b) outer profile code

c) modified scale multiplier

d) scaled coordinates for the inner and outer profiles

e) volume

In the graphics mode a diagram of each pot is drawn on the plotter, in the conventional archaeological style. Note that the only information carried consists of the inner and outer left-hand profiles. The right-hand profile is created by reversing the left-hand profile, with suppression of any reentrant footrim. A vertical centre line and a horizontal rim line are added. The diagrams are labelled.

Similarity coefficients are calculated in the following manner:

1. Treating the 11 percentages of maximum width in the profile code as "coordinates" of the pot in 11-dimensional space, the Euclidean distance is taken between the two pots being compared and then normalised to give the profile coefficient C_p , i.e.

$$C_p = 100 - \sqrt{\frac{\sum_{k=1}^{11} (p_{ik} - p_{jk})^2}{11}}$$

2. The scale ratio $C_s = \left(\text{if } s_i > s_j \text{ then } \frac{100 s_j}{s_i} \text{ else } \frac{100 s_i}{s_j} \right)$

$$\text{and volume ratio } C_v = \left(\text{if } v_i > v_j \text{ then } \frac{100 v_j}{v_i} \text{ else } \frac{100 v_i}{v_j} \right)$$

are also calculated. The volume ratio carries information about the relative heights and widths of the pots, while the scale ratio carries information about the relative heights only. The profile similarity C_p carries information about the relative widths only.

Hence it seems logical to compound the C_p and C_s values to give

$$\text{similarity coefficient } C = \frac{C_p C_s}{100} \quad \text{and this has been done with}$$

sensible results. To indicate that C as defined gives better discrimination between pottery types, Figures 4.6 - 4.8 give scalograms and a dendrogram for C_p alone. These figures should be compared with Figures 1.1, 1.2 and 1.20 respectively, which use C rather than C_p . The program also retains the n best matches between the paper tape record and the magnetic tape records, where n is a specified variable. The program uses the following procedures: OUTSBL (writes tape mark and special block marker), SEARCH (searches for the occurrence of a sequence of characters in any part of a defined string), UNPACK (unpacks a packed string produced by INSTRING to integer characters), READFIG (reads the name of the pot, the scale multiplier, then d-Mac coordinates, rotates, smooths and scales the coordinates to standard height, with adjustment of the scale ratio), DRAWFIG (draws a standardised pot to the specified scale, which may include actual size). Figure 4.9 shows a typical d-Mac tape specifying a pot. Figures 4.10 and 3.14 show typical graph plotter output from this program in the graphics mode, while Figure 4.11 gives a section of line printer output from a typical comparison run.

4.4.4

ARCJW024 Diagram Generator for display unit (ALGOL 60 and EDGAR)

This program, which is fully described in paragraph 3.4.15, may be used to produce pot diagrams for direct publication. The same d-Mac tape as is used to specify a pot for ARCJW019 (see paragraph 4.4.3 above) may be read by this program, normalised, then displayed to actual size at any position on the screen. The scale of the reproduction may be changed by teletype message. The name of the pot and current scale is written below the pot diagram, which is of conventional archaeological type with section to the left and elevation to the right. When the user

MULTIDIMENSIONAL SCALING PLOT

COMMON SAMIAN TYPES

(PROFILE CODES ONLY)

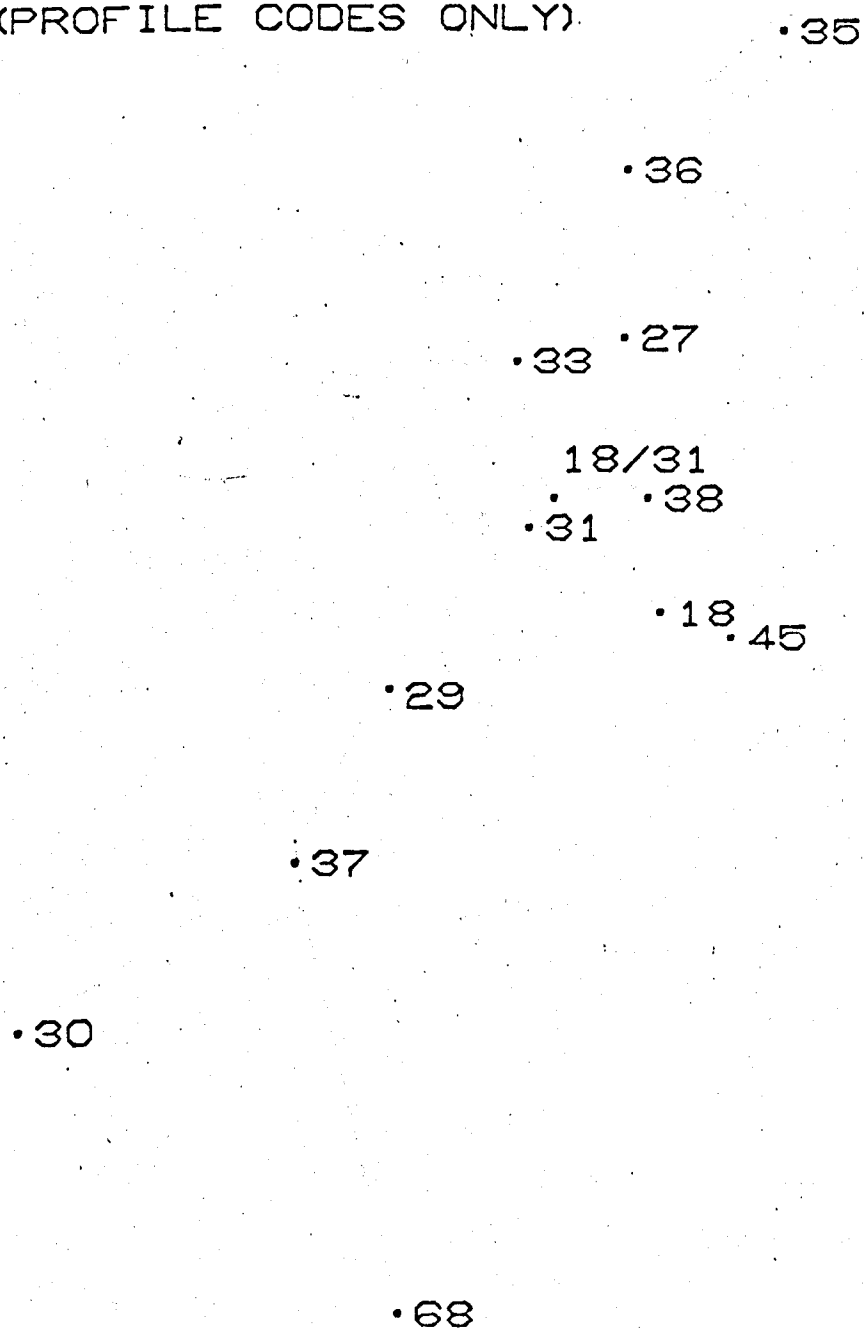


Figure 4.6

COMMON SAMIAN POTTERY TYPES
SCALOGRAM FOR PROFILE CODES ONLY
WITH RENFREW-STERUD LINKAGE

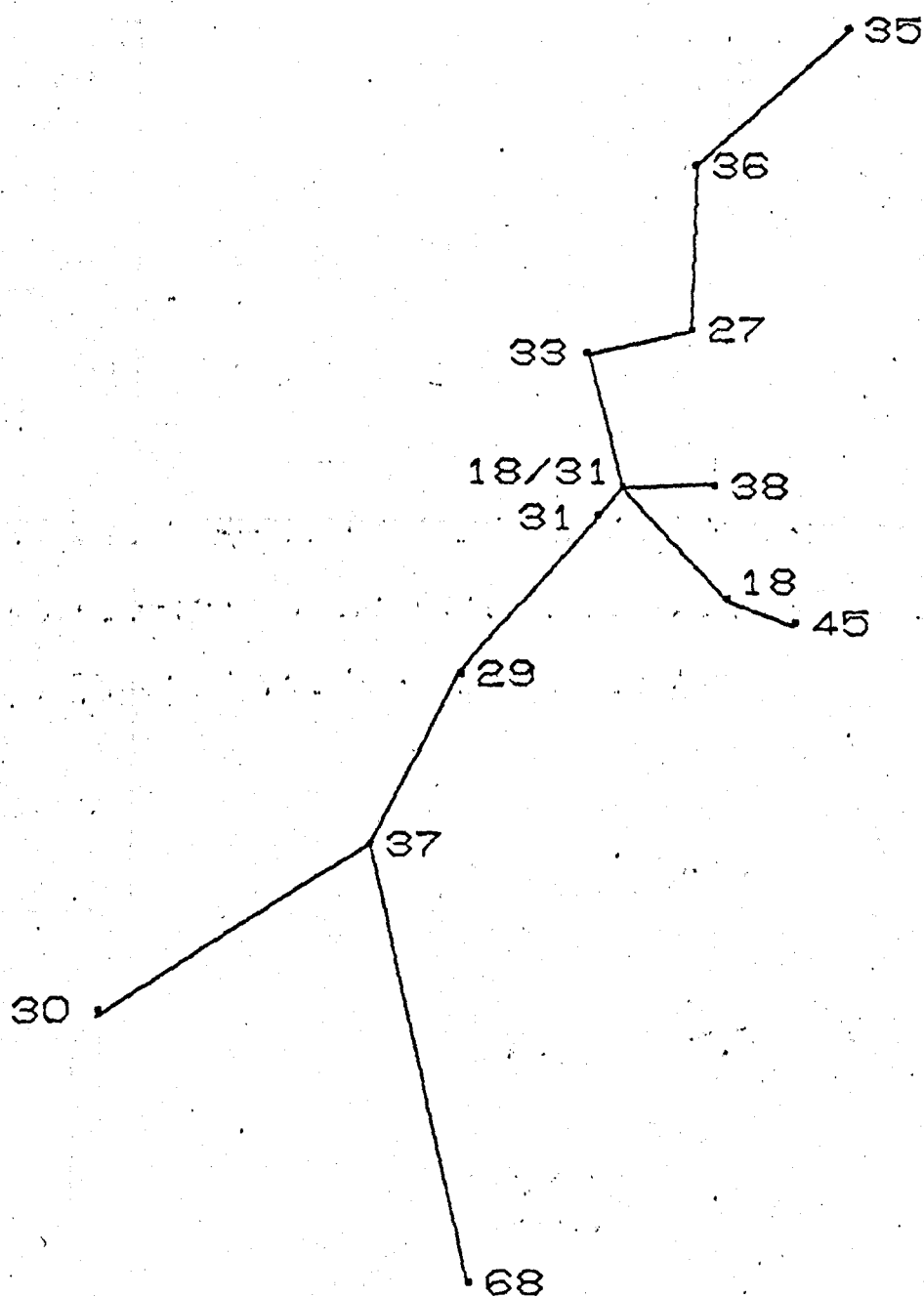
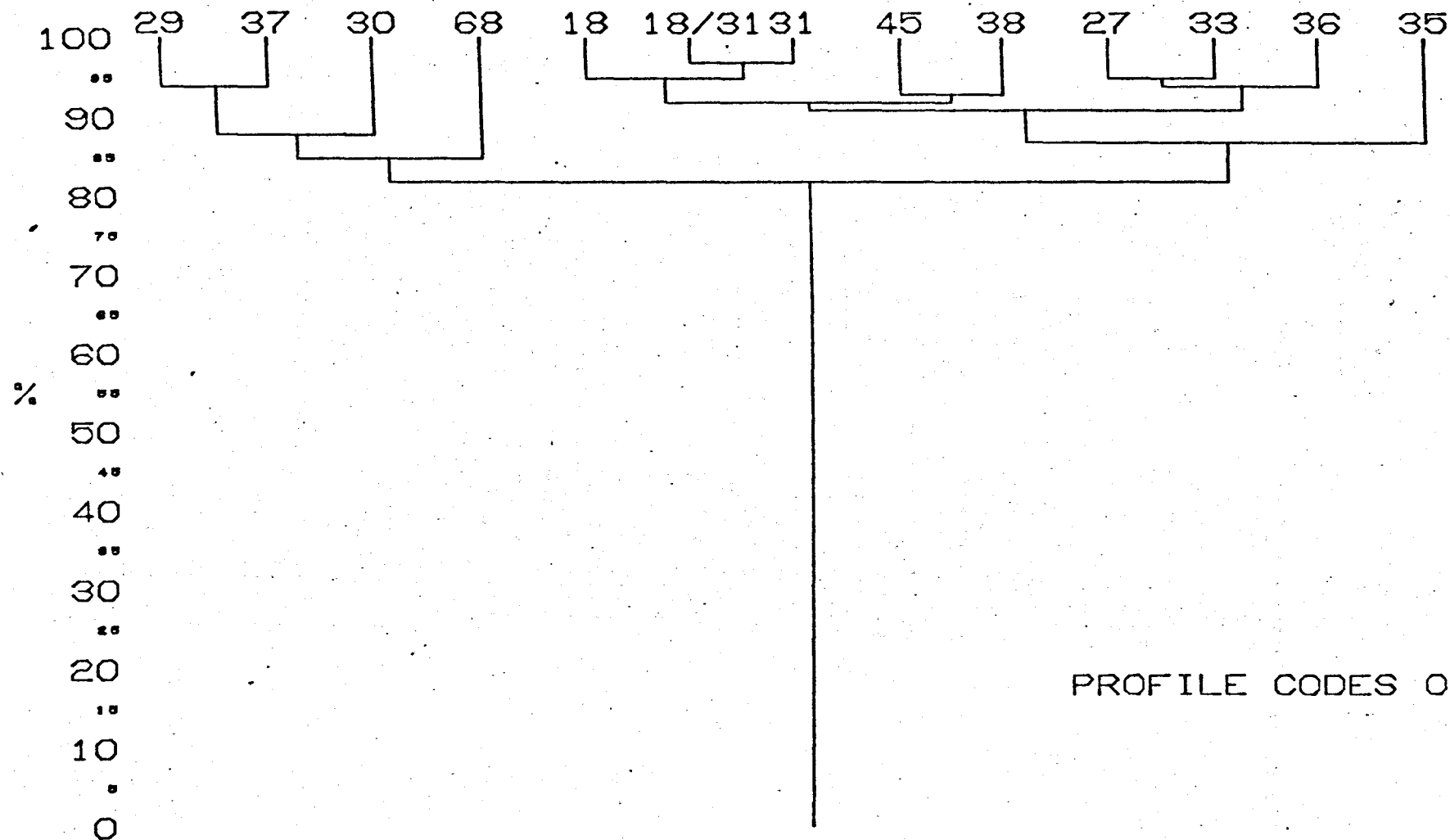


Figure 4.7

PHENON DENDROGRAM FOR COMMON SAMIAN TYPES



PROFILE CODES ONLY

Figure 4.8

*BEAKER S4 9770

3

X3174 Y1854

X3154 Y2489

X0000 Y0000

X3171 Y1909

X3162 Y1907

X3138 Y1904

X3128 Y1903

X3110 Y1897

X3097 Y1897

X3082 Y1891

X3061 Y1893

X3057 Y1895

X3054 Y1899

X3044 Y1912

X3032 Y1929

X3026 Y1945

X3017 Y1963

X3009 Y1985

X3000 Y2001

X2987 Y2033

X2978 Y2059

X2968 Y2085

X2961 Y2105

X2951 Y2141

X2954 Y2157

X2958 Y2176

X2962 Y2194

X2968 Y2221

X2972 Y2238

X2981 Y2283

X2980 Y2295

X2978 Y2327

X2978 Y2351

X2978 Y2369

X2976 Y2385

X2976 Y2403

X2977 Y2424

X2978 Y2437

X2979 Y2457

X2977 Y2467

X2978 Y2472

X2973 Y2477

X2969 Y2481

X2963 Y2483

X2958 Y2483

X0001 Y0001

End of inner profile

Name of pot

Scale multiplier

Alignment points

Start of inner profile

X0000 Y0000

X2960 Y2483

X2950 Y2466

X2948 Y2443

X2951 Y2423

X2948 Y2407

X2952 Y2376

X2952 Y2360

X2951 Y2329

X2951 Y2297

X2950 Y2277

X2950 Y2257

X2947 Y2229

X2938 Y2189

X2937 Y2177

X2929 Y2161

X2929 Y2153

X2931 Y2126

X2940 Y2101

X2946 Y2083

X2949 Y2072

X2961 Y2042

X2967 Y2020

X2981 Y1987

X2994 Y1965

X2999 Y1950

X3006 Y1933

X3016 Y1910

X3029 Y1888

X3033 Y1873

X3037 Y1864

X3043 Y1854

X3048 Y1852

X3048 Y1852

X3048 Y1851

X3052 Y1851

X3062 Y1851

X3068 Y1851

X3081 Y1851

X3086 Y1851

X3099 Y1851

X3119 Y1851

X3136 Y1852

X3146 Y1852

X3162 Y1856

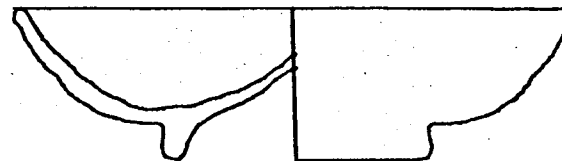
X3176 Y1854

X0001 Y0001

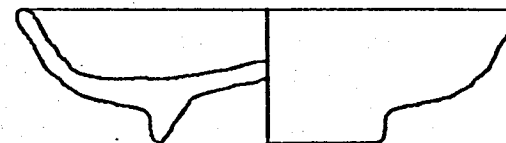
Start of outer profile

End of outer profile

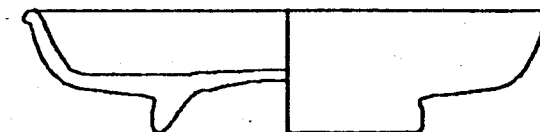
SAMIAN DRAG. 31



SAMIAN DRAG. 18-31



SAMIAN DRAG. 18



Typical graph plotter pot drawings (Samian pottery types)

Figure 4.10

24.94
 1 44.72
 2 71.25
 3 83.76
 4 90.67
 5 95.08
 6 96.35
 7 96.79
 8 96.91
 9 97.47
 10 97.75
 TALL TYPE
 VOLUME 1.27 LITRES

NAME	PROFILE COEFF.	SCALE RATIO	SCALE COEFF.	VOLUME RATIO	VOLUME COEFF.	SIMILARITY COEFF.
SAMIAN DRAG. 29	94.07949	0.87713	87.71331	1.25157	79.89950	82.52023
SAMIAN DRAG. 37	100.00000	1.00000	100.00000	1.00000	100.00000	100.00000
SAMIAN DRAG. 18	86.84121	0.33788	33.78840	0.14544	14.54403	29.34225
SAMIAN DRAG. 18-31	87.58212	0.38225	38.22526	0.13679	13.67925	33.47849
SAMIAN DRAG. 31	88.36404	0.41980	41.97952	0.21777	21.77673	37.09480
SAMIAN DRAG. 30	90.61595	1.69966	58.83534	1.85849	53.80711	53.31420
SAMIAN DRAG. 68	86.43524	1.60068	62.47335	0.81053	81.05346	53.99899
SAMIAN DRAG. 27	83.42624	0.50171	50.17065	0.10220	10.22013	41.85549
SAMIAN DRAG. 33	85.45104	0.47440	47.44027	0.12107	12.10692	40.53821
SAMIAN DRAG. 35	76.65747	0.31741	31.74061	0.03302	3.30189	24.33155
SAMIAN DRAG. 45	88.43908	0.90444	90.44369	0.79245	79.24528	79.98756
SAMIAN DRAG. 36	80.10136	0.33447	33.44710	0.16824	16.82390	26.79158
SAMIAN DRAG. 38	87.56846	0.76109	76.10922	0.62657	62.65723	66.64766

4 HIGHEST SIMILARITY COEFFICIENTS

	SIM.	SCALE	VOLUME
SAMIAN DRAG. 37	100.00000	100.00000	100.00000
SAMIAN DRAG. 29	82.52023	87.71331	79.89950
SAMIAN DRAG. 45	79.98756	90.44369	79.24528
SAMIAN DRAG. 38	66.64766	76.10922	62.65723

Typical line printer output on comparison run

Figure 4.11

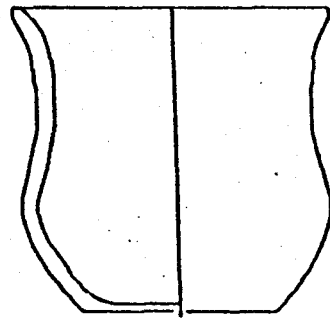
is satisfied with the scale of the reproduction the pot is permanently written into the display file. Several pots may be positioned on the same diagram, and decoration features may be added using the sketching facility. Typical pottery diagrams are shown in Figures 4.12, 4.13 and 3.18.

4.4.5

POTVOL Pottery volumes (BASIC - designed for remote terminal use).

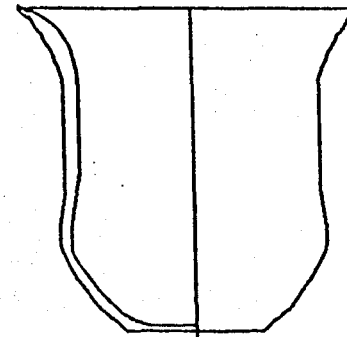
This is a version in BASIC of the algorithm used in ARCJW005 (see paragraph 4.4.1 above). The data consists of a string identifier, scale multiplier and type marker for the pot, plus a series of readings at 2mm intervals of the height (for "tall type" pots) or of the radius (for "dish type" pots). Thus the computer in this case is being used to replace the hand calculator directly - there is no smoothing, rotation and internal scaling as in ARCJW005. The output of the program is the name and type of the pot plus the calculated volume in litres.

COMMON BEAKER POTTERY TYPES: CLARKE CLASSIFICATION



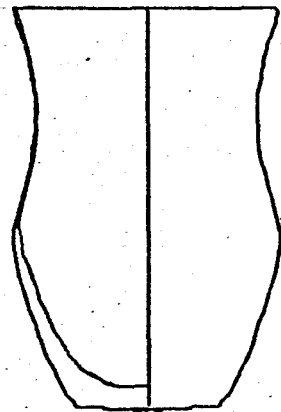
BEAKER AOC 6

SCALE 0.25



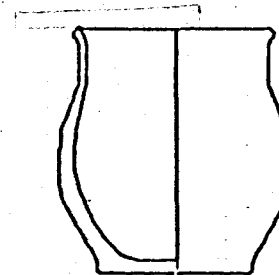
BEAKER E 65

SCALE 0.25



BEAKER N-MR 233

SCALE 0.25

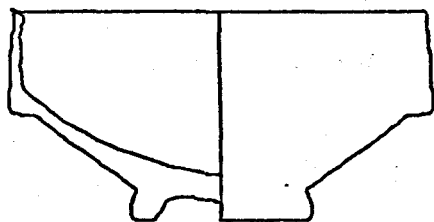


BEAKER N-NR 287

SCALE 0.25

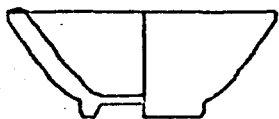
COMMON SAMIAN POTTERY TYPES

Typical graphics diagram generator output
(Samian pottery types)



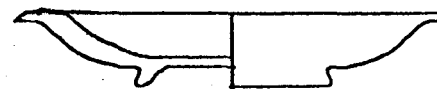
SAMIAN DRAG. 45

SCALE 0.25



SAMIAN DRAG. 33

SCALE 0.25



SAMIAN DRAG. 36

SCALE 0.25



SAMIAN DRAG. 35

SCALE 0.25

Figure 4.13

Pottery Classification Bibliography

- Ankel, C., 1966
Methods of research on prehistoric pottery
by means of computers
*Proceedings of the international symposium
on mathematical and computational methods
in the social sciences*
International Computation Center, Rome
- Ankel, C., 1967
V5. Maschinelle Dokumentation prähistorischer
Keramik
C.Hum.
in *Directory of Scholars Active 1967/68* (1), 93
- Clarke, D.L., 1962
Matrix analysis and archaeology with particu-
lar reference to British Beaker Pottery
Proc. Prehist. Soc. N.S. 28, 371-382
- Clarke, D.L., 1970
Beaker Pottery of G.B. and Ireland
Vol.1, 280 pp. Vol.2, 296 pp.
Cambridge
- Dauterman, C., 1968
Sèvres porcelain and the computer: a pilot
project
*Metropolitan Museum conference on computers
and their potential applications in museums*
New York, April 15-17, 1968
- Freeman, H., 1961
On the encoding of arbitrary geometric
configurations
I.R.E. Trans. on Electronic Computers
June 1961, 260-268
- Freeman, H., 1968
Computer methods for the processing,
classifying and matching of profiles and
other irregular curves
*Metropolitan Museum conference on computers
and their potential applications in museums*
New York, April 15-17, 1968
- Gaines, S.W., 1970
Computer ceramics
Newsletter of Computer Archaeology 5(3), 2-4
- Gardin, J.C., 1967
Methods for the descriptive analysis of
archaeological material
American Antiquity 32(1), 13-30
- Holzhausen, H. and Rottländer, R.C.A., 1970
Standardization of provincial Roman pottery
IV. The origin of standardization
Archaeometry 12(2), 185-195
- Institution of Electrical Engineers, 1968
Pattern recognition
*I.E.E. N.P.L. Conference on pattern
recognition*
I.E.E. Conference publication 42
London 373 pp.

- Irwin,H.T., Hurd,D.J. and LaJeunesse,R.M., 1971
Description and measurement in anthropology
 Computer Series: 1, Laboratory of
 Anthropology,
 Report of Investigations no.48
 Pullman, Washington
- Jardine,N., 1971
 A new approach to pattern recognition
Nature 234, 526-528
- McGimsey,C.R. and Green,D.F., 1965
IBM ceramic code outline
 University of Arkansas Museum
- McPherron,A., 1966
 S21. Study of variability in pottery style
C.Hum.
 in Directory of Scholars Active 1966/67, 235
- Orton,C.R., 1970
 The production of pottery from a Romano-
 British kiln site: a statistical investigation
World Archaeology 1(3), 343-358
- Orton,C.R., 1971
 On the statistical sorting and reconstruction
 of the pottery from a Romano-British kiln
 site
 in Hodson,F.R., Kendall,D.G. and Tautu,P. (eds)
 1971
*Mathematics in the archaeological and
 historical sciences*
 Edinburgh University Press, Edinburgh 453-459
- Perlman,I. and Asaro,F., 1971
 Pottery analysis by neutron activation
 in Brill,R.H. (ed.)
Science and Archaeology 182-195
 M.I.T. press, Cambridge, Massachusetts
 and London 288 pp.
- Rosenfeld,A., 1969a
 Picture processing by computer
Computing Surveys 1(3), 147-176
- Rosenfeld,A., 1969b
Picture processing by computer
 Academic Press, New York and London 196 pp
- Rottländer,R.C.A., 1966
 Is provincial Roman pottery standardized?
Archaeometry 9, 76-91
- Rottländer,R.C.A., 1967
 Standardization of Roman provincial pottery
 II: Function of the decorative collar on
 form Drag.38
Archaeometry 10, 35-45
- Rottländer,R.C.A., 1969
 Standardization of Roman provincial pottery
 III: The average total shrinking rate and
 the bills of La Graufesenque
Archaeometry 11, 159-164
- Rowlett,R., 1968
 Capitaine Leon Berard, Prehistorian (1883-1918)
Antiquity 42, 315-318

- Rutovitz,D., 1966 Pattern Recognition
J. Royal Statistical Society 28(4), 504-530
- Shepard,A.O., 1957 *Ceramics for the archaeologist*
Carnegie Institute of Washington 414 pp.
- Sorenson,J.L., 1964 Use of automated tools in archaeology
American Antiquity 30(2), 205-206
- Stones,C.J. and Williamson,J.B., 1970
Sèvres porcelain and its incised marks
Science and Archaeology 4, 15-18
(project supervised and final paper edited
by Wilcock,J.D.)
- Wagner,N., 1971 Computerised pottery analysis
Newsletter of Computer Archaeology 7(2), 2
- Watanabi,S., 1969 (ed.) *Methodologies of pattern recognition*
Academic Press, New York and London 578 pp.

30 references

Chapter 5

ARCHAEOLOGICAL SURVEY REDUCTION

5. *Archaeological Survey Reduction*

5.1

A particularly obvious application of the computer in archaeology is the routine reduction of site readings taken by geophysical instruments such as the *resistivity meter*, *proton magnetometer*, *proton gradiometer*, *fluxgate gradiometer*, *pulsed magnetic induction meter*, or *Banjo meter*. Following the grid survey of the site it is customary to explore the archaeological potential of the site with these instruments, which can detect archaeological features below ground and give some indication of the most profitable places to dig. A common procedure is to take readings at 1m spacings, thus even a single 100m square produces 10,000 readings, and an archaeological site may be several hectares in area. Also there are usually disturbances due to diurnal changes in the earth's magnetic field, radio transmissions, power cables, ground currents, DC electric trains, metallic litter, geological features and random noise. Drainage of the site, recent rainfall and magnetic storms also affect some types of instruments. This need for the reduction of large numbers of readings, with removal of effects not caused by archaeological features such as foundations, buried pits and ditches, makes the problem an ideal application for the computer. The elimination of non-archaeological effects is called *filtering*, and the computer may also plot a map of the site using symbols or dots of various densities to indicate the most profitable places to dig.

5.2

This section will review the various types of instrument available and the application of the computer to the problem by various workers.

5.2.1

When using resistivity meters it is necessary to insert electric current into the ground by some electrodes and detect the resulting potential between others, thus measuring the *resistivity* of the ground. High resistivity is experienced over buried foundations. Various electrode configurations have been devised for this application, the most famous of which was first described by Wenner (1916). In the *Wenner configuration* the 4 electrodes are placed in a straight line with equal spacings. The current enters the ground via the outer pair of electrodes, while the inner pair is used for measuring the potential. The actual survey point is taken as being mid-way between the inner pair of electrodes, and the resulting penetration of the current (and hence depth of the survey) is approximately equal to the electrode spacing. The electrodes may be advanced in either the *end-on* or *broadside* mode. In the end-on mode only one electrode needs to be advanced at a time, a single electrode taking the functions of right-hand current electrode, right-hand potential electrode, left-hand potential electrode and left-hand current electrode (assuming the traverse is advancing to the right). This is the most frequently employed mode since it entails the repositioning of only one electrode at a time, but it has the disadvantage of the "*double-peak*" effect. A high-resistance feature will be indicated by a high reading as the first current electrode crosses the feature, a lower reading as the potential electrodes cross, and a second high reading as the second current electrode crosses. The actual location of this high resistance feature is mid-way between the two high peaks, thus the interpretation of an end-on traverse survey requires some experience. Alternatively the broadside mode may be employed where a single peak

is obtained, at the correct location for a high resistance feature, but it has the disadvantage that all four electrodes must be moved for every new reading.

5.2.2

Tagg (1956,1957) describes a geophysical survey in the Wash area using the Megger Earth Tester Mk I manufactured by his employers, Evershed and Vignoles Ltd. The survey had the object of locating the causeway used by King John's pack train carrying his lost treasure. The causeway was located with some success, but of course the chance of locating the actual treasure would be slight. The Megger Mk I was loaned with a transporting frame for the Bull Pot Farm, Westmorland preparations for the Oxford University Expedition to Northern Spain 1961 (see Wilcock, 1963, 1965, paragraph 5.2.12 below).

5.2.3

Dean (1958) describes a method for the synthesis of filter transfer functions using frequency analysis. The method is applicable to gravity and magnetic surveys, and has been elaborated by Scollar (see paragraph 5.2.10 below).

5.2.4

Palmer (1959,1960) was one of the first workers to employ resistivity surveys for archaeological purposes, and he designed a new probe configuration, in which each potential electrode is closely linked with its nearest current electrode by a wooden rod. The two pairs of electrodes are then moved independently and the potential probe spacing is typically about 80% of the current probe spacing. Because the potential probe spacing is a variable proportion of the current probe spacing, reduction formulae become necessary for the calculation of the resistivity, and these are given by Palmer.

Advice was given by Palmer at Wells Museum shortly before his death to the Oxford University Expedition to Northern Spain, 1961 (see Wilcock, 1963, 1965, paragraph 5.2.12 below).

5.2.5

Dr Martin Aitken (1959,1960,1961) of the Research Laboratory for Archaeology and the History of Art, Oxford was one of the first workers to systematically explore the possibilities of different types of instrument in archaeology. He continued to use resistivity meters, but also developed magnetic instruments in collaboration with Dr E.T.Hall, Director of the laboratory and Director of Littlemore Scientific Engineering Co.Ltd, a subsidiary largely financed by Hall's Breweries set up to manufacture instruments for archaeological and other purposes. Aitken's 1959 paper describes the *proton magnetometer*, an absolute instrument which measures the earth's magnetic field to an accuracy of nearly one part in 50,000 using *proton free-precession*. The detecting element is a "bottle" consisting of a large coil of wire wound around a polythene bottle containing a supply of protons in the form of distilled water or methyl alcohol. The instrument first passes a strong current through the coil during the *polarising period*, creating a strong axial magnetic field which causes a proportion of the protons, which have elementary magnetic moments, to align along the axis of the bottle, normally set up to be magnetic east-west. When the current is removed during the *precession period* the protons move back towards the direction of the earth's magnetic field, magnetic south-north. However, because the protons also have an elementary spin the *gyroscopic effect* operates and the protons move back in a conical, "spinning-top" locus called *precession*. As they do this the elementary magnetic moments of the protons induce an oscillating current in the coil of wire. The

frequency of precession, and hence the frequency of the oscillating current is a function of the earth's magnetic field strength (possibly modified by archaeological remains) at the bottle. Thus by using pulse shaping and counting circuits it is possible to measure the earth's magnetic field to a high degree of accuracy. Typical features which modify the earth's magnetic field on an archaeological site are buried pits (which are important for they carry most of the dating material for a site) and especially *thermo-remanent* features such as pottery kilns and hearths (these have been fired in antiquity and have "locked-in" the direction of the earth's magnetic field into their ferromagnetic constituents at the time of their last firing, hence causing a big anomaly at the present day).

Although the proton magnetometer will detect these important features, it suffers from the disadvantage that the earth's magnetic field exhibits a diurnal change, and check readings must be taken frequently at a reference location and subtracted from the other readings if they are to have any relative significance over a large site which takes several hours to survey. For this reason the *proton gradiometer*, described in Aitken's 1960 paper, was developed. This uses two bottles, one held high in the air on a staff, detecting the average earth's magnetic field, and a second close to the ground, detecting the earth's magnetic field modified by the archaeological features. The two readings are subtracted in the electronics, just leaving the archaeological effects. Since the diurnal change effects both bottles equally, the subtraction removes it entirely and no reference readings need to be taken.

Aitken's 1961 book, which is still useful although now inevitably dated, is a reference work covering the proton magnetometer and gradiometer, resistivity surveying, general dating methods,

radiocarbon dating, magnetic dating and physical and chemical analysis of finds.

5.2.6

Nash and Thompson Ltd. (1960) give operating instructions for the Tellohm range of soil resistance meters. These are much more powerful than the usual transistorised meters used for archaeological purposes (see paragraph 5.2.8 below), and hence penetrate deeper into the ground; they are about the size of a suitcase, operate from dry batteries, have a direct-reading dial and do not suffer from the hand-cranking disadvantage of the Megger range of instruments.

5.2.7

Evershed and Vignoles Ltd (1961) in two publications give resistivity techniques for the Megger range of instruments.

5.2.8

Martin and Clark (1961) give operating instructions for the transistorised instrument which became commercially available in 1960 and is now used by many archaeologists for resistivity surveying. It is small, light, operates from a flat torch battery and is of fairly low power, but adequate for most archaeological sites. A useful feature is the rotating switch which allows four electrodes to be operable while a fifth is free for moving to the next electrode position in an end-on traverse. The switch is then rotated one position, engaging the newly-positioned electrode as a current electrode and freeing the electrode at the rear for movement. Clark now works for the Department of the Environment in the geophysical division responsible for surveying ancient monuments.

5.2.9

Tite compares various instruments at the Iron Age fort at Rainsborough, near Banbury, Oxfordshire (Tite, 1961; Aitken and Tite, 1962). The instruments used were the proton magnetometer, proton gradiometer

and fluxgate gradiometer.

5.2.10

The major worker in the automation of archaeological surveys is Irwin Scollar, of the Rheinisches Landesmuseum, Bonn. His work is reported in numerous papers (Scollar, 1962, 1965a, 1965b, 1966, 1967, 1968a, 1968b, 1969a, 1969b, 1970a, 1970b; Scollar and Krückeberg, 1966, 1969; Scollar and Linington, 1966; Black and Scollar, 1969; Gubbins, Scollar and Wisskirchen, 1971). The 1962 paper is a theoretical study of the basis of magnetic surveying. Calculations are carried out by Krückeberg on the IBM 1410 machine. Scollar has designed and built an automatic survey station in a minibus, including magnetometer electronics, digitiser, paper tape punch and tape recorder to give commands to the operator moving the bottle staff. The paper tape is later read by computer and analysed. Scollar's 1966 paper describes in detail the soil of the Rhineland. Some simple FORTRAN programs are also described. Scollar and Krückeberg have developed complex filtering to improve the contrast in dot-density plots of magnetometer survey data. Flowcharts, symbol diagrams and dot plots are given. The software is in FORTRAN for the IBM 410/7090 and IBM 360/50 machines fitted with magnetic tapes, discs, Rand tablet and CRT with light pen. An off-line plotter is also used. Scollar has also collaborated with Richard Linington of the Fondazione Lerici, Rome in the computer reduction of proton gradiometer readings in Etruria. In 1969 Scollar describes a simulation program with display output, written for the IBM 7090 (a development of the work of Linington) and some techniques for filtering. In the development of dot-plots, filtering is used to remove overall background noise. Various types of filter are described:

- (a) subtraction of the peripheral average from the central value (high pass);
- (b) addition of the peripheral average (low pass);
- (c) combination of peripheral circles with weighted averages of various signs (band-pass).

Examples of simulated data exhibiting various characteristics are given. Random noise is added, and filtering used to restore the picture. The best technique is found to be *point source removal* (replacement of a wild value by the average of peripheral points taken somewhat away from the central point, if this wild value deviates by an absolute value greater than a fixed ratio from the average value, *low-pass filter*, then *high-pass filter*. A discrimination level is also found to be useful in removing all readings below a certain value. Multiplication of readings gives enhanced blackness for dot-plots on the graph plotter. Black and Scollar describe the use of the Cooley-Tukey fast Fourier transform for filtering. In 1970 Scollar describes further filtering methods, including the two-dimensional fast Fourier transform. Isometric plots of the spectra of a Roman camp are given with simulated soil noise using both equal probability and Gaussian probability vertical dipole distributions. A pseudo-perspective plot of the magnetic field of a circular ditch is given, with its Fourier transform. Typical filters are normalised spectra of ring ditches or stylised Roman camps (in this case the feature must be rotated to match the archaeological remains). Gubbins, Scollar and Wisskirchen give a general theory for the filtering of multi-dimensional data. This theory is applied to all orthogonal functions and Haar and Walsh transforms are found to be useful for quick computation. Algorithms are given for these functions.

Two-dimensional data is treated extensively, with the description of several types of near-optimum filters using known properties of the data. Magnetic prospecting data with a very poor *signal to noise ratio* is used as an example.

5.2.11

R.J.C. Atkinson (1963) describes the use of the Megger Earth Tester in archaeology, which he first used at Dorchester (Oxfordshire) in 1946. A circuit is given for a switching system which operates in conjunction with five probes (cf. the Martin-Clark system described in paragraph 5.2.8 above). The system may be assembled from standard wafer switch components.

5.2.12

A selection of resistivity meters were used on the Oxford University Expedition to Northern Spain 1961 (Wilcock, 1963, 1965), which was concerned with both archaeology (in particular petroglyphs) and speleology. The expedition was supported and advised by Tagg (see paragraph 5.2.2), Palmer (paragraph 5.2.4), Aitken (paragraph 5.2.5), Nash and Thompson Ltd. (paragraph 5.2.6), Evershed and Vignoles Ltd (paragraph 5.2.7) and Martin and Clark (paragraph 5.2.8). The resistivity meters used were the Megger Mk I, the Nash and Thompson Geophysical Tellohm and the Martin-Clark Resistivity Meter. A special pack frame was constructed to carry four 300 ft reels of wire. A slide-rule method was developed, based on Palmer's theory for the calculation of resistivity using various ratios of potential to current electrode spacing. The Wenner methods of end-on, broadside and expanding electrode arrangements were used. The 1965 publication gives appendices on the theory, the slide-rule method, and Tellohm technical data.

5.2.13

J.C. Alldred (1964), of the Research Laboratory for Archaeology

and the History of Art, Oxford gives drawings and circuitry for a *fluxgate gradiometer* and the principles of its operation. The fluxgate elements must be rigidly aligned and this means that a bulky cylindrical container must be carried around the site by two operators. For these reasons the fluxgate gradiometer is not used a great deal.

5.2.14

E.K.Ralph (1964), Associate Director of MASCA, compares a proton gradiometer and a rubidium gradiometer for archaeological prospecting. The Varian rubidium gradiometer described is much more sensitive than the proton gradiometer. Work on the site of Sybaris is described.

5.2.15

Filtering of geophysical readings of poor signal to noise ratio was made much easier by the development of the fast Fourier transform (Cooley and Tukey, 1965). This is an algorithm to perform digital filtering by the machine calculation of complex Fourier series. It is used extensively by Scollar (see paragraph 5.2.10 above).

5.2.16

Tagg (1965) describes a method of constructing 3D resistivity maps. Resistivity curves are pasted on thin card, cut out, then slotted together at rightangles into a 3D "egg box" grid. Tagg also simulates resistivity anomalies using nylon blocks (high resistivity) to represent archaeological features in a water tank (low resistivity) representing the soil. The depth of water was $1\frac{1}{4}$ inch, the nylon blocks were of 1 inch square cross-section and of various lengths, and an electrode spacing of 1 inch was used.

5.2.17

Wilcock and Gee (1965) describe the tracing of a Roman road using a Martin-Clark resistivity meter.

5.2.18

The *pulsed magnetic induction meter* is described by Colani (1966), and Colani and Aitken (1966). This meter gives results similar to those of the proton gradiometer, and in addition detects metal very well. The meter transmits into the ground. The receiver is disabled during the transmit pulse, then switched on for a limited time, when it receives magnetic fields caused by induction in metallic items and such archaeological features as buried pits and ditches. The sites surveyed with the instrument, Waddon Hill (where pits were detected), Woodeaton Roman temple site (metallic votive offerings) and Blewburton (Saxon burials with metallic grave goods within an Iron Age hill fort), demonstrate that the results are similar in some respects to those obtained by the proton gradiometer.

5.2.19

Mark Howell (1966) describes the "*Banjo*" soil conductivity meter, so called because of its shape. The results are similar to those obtained by the pulsed magnetic induction meter. A transmitter is mounted in a box at the rear of a staff which is carried horizontally, and a circular receiver coil is mounted at the forward end. The meter has been used at South Cadbury and the results analysed by computer (Wilcock, see paragraph 5.2.22 below). Since this time Howell has developed a dot-plotting machine, which plots a magnetic map using a drum and felt-tip pen mounted in a pen carriage, controlled by tape-recorded readings.

5.2.20

Wilcock (1966) describes the use of the Martin-Clark resistivity meter at Lawn Farm (where buried masonry of a manor house were detected) and at Trent Vale Roman pottery site. Both sites are near Stoke-on-Trent.

5.2.21

The excavation of South Cadbury Castle, an Iron Age hill fort conquered by the Romans and with Dark Age and Saxon fortifications, the supposed site of "Camelot", is described by the Site Director, Leslie Alcock (1967a,1967b,1968a,1968b,1969a,1969b,1970a,1970b,1971,1972) and by Alcock, Musson and Evans (1967 onwards). Geophysical instruments have been used systematically to indicate the most profitable places to dig. The site is about 7 hectares in area, and a computer was called into service (ICL KDF9 at Kids Grove, Staffordshire, see Wilcock, paragraph 5.2.22 below).

5.2.22

C.R. Musson (1968), Deputy Site Director of the South Cadbury site, describes the geophysical survey carried out by Howell and Aitken, the readings of which were made available for computer processing by ICL KDF9 at Kids Grove, Staffordshire and the results transmitted by Telex. This operation, probably the first transmission of computer-processed archaeological data in Britain is described by Wilcock in three popular press articles (1968a,1968b,1968c) and in four other papers (1969a,1969b,1970a,1970b). The design of a field data form for the entry of geophysical readings is described in detail. Dot-density plots have been prepared by computer (see Figure 3.5). At first it was hoped that random dots could be positioned within circles of radius $a\sqrt{2}/2$ centred on the reading points which have spacing a , thus representing more accurately

the detection range of the Banjo meter (see paragraph 5.2.19 above), but unfortunately the overlap of adjacent circles produces more heavily stippled areas which give the impression of regular observation points. Although the heavier areas are centred between four reading points, the effect is to be avoided, since the final dot-density plot should not give any indication of the pattern of readings. Accordingly the dots are positioned within square regions of side a , centred on the reading points. This development is described in the 1970 *Prospezioni Archeologiche* paper by Wilcock.

5.2.23

Richard Linington, of the Fondazione Lerici, Rome has been a major worker in the use of geophysical instruments, and the use of computers (in collaboration with Scollar) for the reduction of readings. The journal *Prospezioni Archeologiche* is published by the Fondazione Lerici as a vehicle for papers on geophysical topics. Linington describes his work in several papers (Linington, 1967, 1968a, 1968b, 1968c, 1969, 1970a, 1970b; Scollar and Linington, 1966). A survey of La Civita, Tarquinia (Etruscan tomb site) is described, the readings from which were processed by Scollar and Krückeberg. Linington reports FORTRAN IV and assembler language software for the IBM 360/40 machine for the treatment and representation of archaeological prospecting results. This is a similar project to the work of Scollar at Bonn, but it has been developed independently at Rome. The survey results are punched on cards or paper tape, checked by computer, then transferred to magnetic tape. Where necessary, mathematical treatment is applied to the results and then symbol diagrams are produced on the line printer (no plotter was available). Specially-designed forms are used for the recording

of data in the field and for direct data preparation in the computer bureau. Wilcock also describes a similar form (see paragraph 5.2.22 above). The software allows correction for diurnal variations, filtering, and representation of the results by graphs, isometric diagrams, contour diagrams and symbol diagrams. Linington describes a simple filtering method, requiring only a small amount of computer time, which seems to yield interesting results.

5.2.24

A new type of electrode configuration, the *square array* was designed by A.J.Clark (see paragraph 5.2.8 above) in 1968. The electrodes form the legs of a table which is used to carry the resistivity meter and pad of survey forms. The table is easy to move and if necessary the whole survey can be carried out by one person, whereas the conventional end-on traverse using individual electrodes requires a highly-coordinated team of operators. The contacts to the resistivity meter need not be moved between readings. A suitable form, which has been designed and used on the square array by Wilcock (see paragraph 5.2.22), may be submitted direct to a computer bureau for data preparation. In a 1969 paper Clark describes the Wenner, Palmer and Schlumberger electrode configurations, their advantages and disadvantages.

5.2.25

Aitken (1969) presents comparative surveys carried out by proton magnetometer, proton gradiometer and fluxgate gradiometer.

5.2.26

Breiner (1970) reports a cesium magnetometer survey at San Lorenzo, Mexico with computer processing of the results. The data is input on punched cards, and a map is output on the line printer, with "contours" of equal magnetic intensity. Alternative output is by

electrostatic printer, a device where the black areas receive electrostatic charges to which ink particles adhere and are later fused by a heat process. These electrostatic maps may be produced up to 20 times faster than maps plotted on the digital incremental plotter. The maps are claimed to be superior to hand-drawn maps.

5.3 *Survey Reduction Programs*

5.3.1

ARCJWO01 Archaeological Plotting Generator (ALGOL 60)

This program carries out simple filtering designed for the Banjo meter, then produces dot-density plots (see Figure 3.5). Design criteria for this program are described in Wilcock (1970, *Prospezioni Archeologiche* paper). The program is fully described in paragraph

3.4.1.

5.3.2

RESMAP Archaeological Survey Reduction (BASIC - designed for remote terminal use)

This program analyses survey readings and prints a symbol map for each square.

First a string identifier for a square is read, then a set of eleven traverses, each of eleven readings (to allow overlap on all edges of the square). Each traverse is terminated by a special "end of traverse" marker, and the eleven traverses are followed by an "end of square" marker. These markers are checked to prevent data registration errors.

The maximum and minimum readings and the range are found. A simple filtering process is carried out to divide the range into six regions using the expression:

$$p_i = \text{INT} (5(p_i - \text{min})^2 / (\text{range})^2) + 1$$

where INT is the lower integer.

Thus for a range of 20 and minimum of 15

$$p_i = \text{INT} \left((p_i - 15)^2 / 80 \right) + 1$$

Allocated symbol

15.00	≤	p_i	<	23.94	lies in region 1	blank
23.94	≤	p_i	<	27.65	lies in region 2	.
27.65	≤	p_i	<	30.49	lies in region 3	+
30.49	≤	p_i	<	32.89	lies in region 4	X
32.89	≤	p_i	<	35.00	lies in region 5	*
		p_i	=	35.00	lies in region 6	*

This filter has the property of ignoring low readings and enhancing high readings.

The square identifier, maximum, minimum and range are printed out, followed by the symbol map for the square, each reading being allocated its appropriate symbol. Unfortunately the square is elongated by the teletype, since the character spacing is 1/10" and the line spacing 1/6". This is a common failing of symbol maps. Finally a key is printed giving the range of readings for each symbol type. This program is useful in obtaining symbol maps while actually on site if a teletype is available in the site office or nearby building, with acoustic coupler, telephone and power supply. A typical output square is shown in Figure 5.1.

Survey Reduction Bibliography

Aitken, M.J., 1959

Magnetic prospecting: IV - The proton magnetometer
Archaeology 7, 49-52

Aitken, M.J., 1960

Magnetic prospecting: The proton gradiometer
Archaeology 7, 19-20

Aitken, M.J., 1961

RUN SANDON SQUARE A
New York and London 181 pp.

Aitken, M.J., 1969

MAXIMUM 34
MINIMUM 15
RANGE 19
Thames Valley University, London 681-694

Aitken, M.J. and Fite, M.

0 0
.*.
+... ++
..X.
*.X.
.. ..
..++ ..++
.
..
0 0

Alcock, L., 1967a

..++ ..++
.
..
0 0

Alcock, L., 1967b

..
.
0 0

Alcock, L., 1968a

KEY
15 - 23
. 23 - 27
+ 27 - 29
X 29 - 31
* 31 - 34

Alcock, L., 1968b

READY
15 - 23
. 23 - 27
+ 27 - 29
X 29 - 31
* 31 - 34

Alcock, L., 1969a

READY
15 - 23
. 23 - 27
+ 27 - 29
X 29 - 31
* 31 - 34

Alcock, L., 1969b

READY
15 - 23
. 23 - 27
+ 27 - 29
X 29 - 31
* 31 - 34

Alcock, L., 1970a

READY
15 - 23
. 23 - 27
+ 27 - 29
X 29 - 31
* 31 - 34

Alcock, L., 1970b

READY
15 - 23
. 23 - 27
+ 27 - 29
X 29 - 31
* 31 - 34

Figure 5.1

Typical output square for geophysical site readings
using RESMAP program

Alcock, L., 1972

South Cadbury excavations, 1966-70
Archaeology 10, 1-7

Survey Reduction Bibliography

- Aitken, M.J., 1959 Magnetic prospecting: IV - The proton magnetometer
Archaeometry 2, 40-42
- Aitken, M.J., 1960 Magnetic prospecting: The proton gradiometer
Archaeometry 3, 38-40
- Aitken, M.J., 1961 *Physics and Archaeology*
Interscience, New York and London 181 pp.
- Aitken, M.J., 1969 Magnetic location
in Brothwell, D. and Higgs, E. (eds), 1969
Science in Archaeology
Thames and Hudson, London 681-694
- Aitken, M.J. and Tite, M.S., 1962 Proton magnetometer surveying on some
British hillforts
Archaeometry 5, 126-134
- Alcock, L., 1967a By South Cadbury is that Camelot
Antiquity 41, 50-53
- Alcock, L., 1967b A reconnaissance excavation at South
Cadbury Castle, Somerset, 1966
Antiquaries Journal 47, 70-76
- Alcock, L., 1968a Cadbury Castle, 1967
Antiquity 42, 47-51
- Alcock, L., 1968b Excavations at South Cadbury Castle, 1967:
A summary report
Antiquaries Journal 48, 6-17
- Alcock, L., 1969a South Cadbury excavations, 1968
Antiquity 43, 52-56
- Alcock, L., 1969b Excavations at South Cadbury Castle, 1968:
A summary report
Antiquaries Journal 49, 30-40
- Alcock, L., 1970a South Cadbury excavations, 1969
Antiquity 44, 46-49
- Alcock, L., 1970b Excavations at South Cadbury Castle, 1969:
A summary report
Antiquaries Journal 50, 14-25
- Alcock, L., 1971 Excavations at South Cadbury Castle, 1970:
A summary report
Antiquaries Journal 51, 1-7
- Alcock, L., 1972 Excavations at Cadbury-Camelot, 1966-70
Antiquity 46, 29-38

- Alcock, L., Musson, C.R. and Evans, D.M., 1967 onwards
South Cadbury Information Sheets
 Dept. of Archaeology, University College
 of South Wales and Monmouthshire, Cardiff
- Allred, J.C., 1964 A fluxgate gradiometer for archaeological
 surveying
Archaeometry 7, 14-19
- Atkinson, R.J.C., 1963 in Pyddoke, E. (ed.)
The Scientist and Archaeology
 London, Phoenix House 1-30
- Black, D.I. and Scollar, I., 1969
 Spatial filtering in the wave-vector domain
Geophysics 34(6), 916-923
- Breiner, S., 1970 Cesium magnetometer survey, San Lorenzo,
 Mexico
MASCA Newsletter 6(1), 3
- Clark, A.J., 1968 A square array for resistivity surveying
Prospezioni Archeologiche 3, 111-114
- Clark, A.J., 1969 Resistivity surveying
 in Brothwell, D. and Higgs, E. (eds), 1969
Science in Archaeology
 Thames and Hudson, London 695-707
- Colani, C., 1966 A new type of locating device
 I - The instrument
Archaeometry 9, 3-8
- Colani, C. and Aitken, M.J., 1966
 A new type of locating device
 II - Field trials
Archaeometry 9, 9-19
- Cooley, J.W. and Tukey, J.W., 1965
 An algorithm for the machine calculation
 of complex Fourier series
Math. of Computation 19, 297-301
- Dean, W.C., 1958 Frequency analysis for gravity and magnetic
 interpretation
Geophysics 23, 97-127
- Evershed and Vignoles, 1961a
A pocket book on resistivity prospecting
 Evershed and Vignoles Ltd., London 46 pp.
- Evershed and Vignoles, 1961b
*Resistivity prospecting with Megger Earth
 Testers*
 Evershed and Vignoles Ltd., London
 Publication 245/2

- Gubbins,D., Scollar,I. and Wisskirchen,P., 1971
Two dimensional digital filtering with HAAR
and WALSH transforms
Ann. Geophys. 27(2), 85-104
- Howell,M., 1966
A soil conductivity meter
Archaeometry 9, 20-23
- Linington,R.E., 1967
Magnetic Survey at La Civita, Tarquinia
Prospezioni Archeologiche 2, 87-89
- Linington,R.E., 1968a
S51. Treatment and representation of archaeological prospecting results
C.Hum. 2(5)
in Directory of Scholars Active, 1968 248
- Linington,R.E., 1968b
S51. Treatment and representation of archaeological prospecting results
C.Hum.
in Directory of Scholars Active, 1968/69(1), 315
- Linington,R.E., 1968c
The Rome computer system for treating archaeological survey results: First part
Prospezioni Archeologiche 3, 19-36
- Linington,R.E., 1969
The Rome computer system for treating archaeological survey results: Second part
Prospezioni Archeologiche 4, 9-58
- Linington,R.E., 1970a
A first use of linear filtering techniques on archaeological prospecting results
Prospezioni Archeologiche 5, 43-54
- Linington,R.E., 1970b
Techniques used in archaeological field surveys
Phil. Trans. Roy. Soc. Lond. A269, 89-108
- Martin and Clark,A.J., 1961
Operating instructions for the Martin-Clark Resistivity Meter
Martin and Clark, Farnham
- Musson,C.R., 1968
A geophysical survey at South Cadbury Castle, Somerset, using the Howell soil conductivity anomaly detector (SCM)
Prospezioni Archeologiche 3, 115-121
- Nash and Thompson, 1960
Operating instructions for the Tellohm range of soil resistance meters
Nash and Thompson, Ltd., Chessington
- Palmer,L.S., 1959
Examples of geoelectric surveys
Proc. I.E.E. 106A, 231-244
- Palmer,L.S., 1960
Geoelectric surveying of archaeological sites
Proc. Prehist. Soc. 26, 64-75

- Ralph, E.K., 1964
Comparison of a proton and a rubidium magnetometer for archaeological prospecting
Archaeometry 7, 20-27
- Scollar, I., 1962
Electromagnetic prospecting methods in archaeology
Archaeometry 5, 146-152
- Scollar, I., 1965a
A contribution to magnetic prospecting in archaeology
Archaeo-Physika, Band 15 der Beihefte der *Bonner Jahrbücher* 21-92
- Scollar, I., 1965b
Archaeological prospecting instruments: Magnetic detection
MASCA Newsletter 1(1), 1
- Scollar, I., 1966
Recent developments in magnetic prospecting in the Rhineland
Prospezioni Archeologiche 1, 43-52
- Scollar, I., 1967
Computer evaluation of magnetic surveys
MASCA Newsletter 3(1), 1
- Scollar, I., 1968a
A program package for the interpretation of magnetometer data
Prospezioni Archeologiche 3, 9-18
- Scollar, I., 1968b
Automatic recording of magnetometer data in the field
Prospezioni Archeologiche 3, 105-110
- Scollar, I., 1969a
Some techniques for the evaluation of archaeological magnetometer surveys
World Archaeology 1(1), 77-89
- Scollar, I., 1969b
A program for the simulation of magnetic anomalies of archaeological origin in a computer
Prospezioni Archeologiche 4, 59-83
- Scollar, I., 1970a
Fourier transform methods for the evaluation of magnetic maps
Prospezioni Archeologiche 5, 9-42
- Scollar, I., 1970b
Magnetic methods of archaeological prospecting - advances in instrumentation and evaluation techniques
Phil. Trans. Roy. Soc. Lond. A269, 109-119
- Scollar, I. and Kruckeberg, F., 1966
Computer treatment of magnetic measurements from archaeological sites
Archaeometry 9, 61-71
- Scollar, I. and Kruckeberg, F., 1969
S87. Data processing of geophysical measurements made on archaeological sites
C.Hum.
in *Directory of Scholars Active 1968/69*(2), 138

- Scollar, I., and Linington, R.E., 1966
Data processing of geophysical measurements on archaeological sites
Proceedings of the international symposium on mathematical and computational methods in the social sciences
International Computation Center, Rome
- Tagg, G.F., 1956
Megger Earth Tester used in search for King John's treasure
Evershed News 4(5), 3-9 and 12
Evershed and Vignoles, Ltd., London
- Tagg, G.F., 1957
A resistivity survey in the Wash area
Journal I.E.E. 3,5
- Tagg, G.F., 1965
Three dimensional resistivity maps
Evershed News 8(2), 10-13
Evershed and Vignoles, Ltd., London
- Tite, M.S., 1961
Alternative instruments for magnetic surveying:
Comparative tests at the Iron Age hill-fort at Rainsborough
Archaeometry 4, 85-90
- Wenner, F., 1916
Wenner configuration of electrodes
Bulletin of the U.S. Bureau of Standards 12, 469
- Wilcock, J.D., 1963
Cave Surveying in Northern Spain
Evershed News 7(7), 12-14
Evershed and Vignoles, Ltd., London
- Wilcock, J.D., 1965
Oxford University Expedition to Northern Spain 1961: Geophysical survey and results
Cave Research Group of Great Britain Publication 14, 31-38
- Wilcock, J.D., 1966
Scientific aids review
City of Stoke-on-Trent Museum Archaeological Society Reports 2, 11-14
- Wilcock, J.D., 1968a
Computer at court of King Arthur (Press report)
Daily Mirror, August 15, 1968, 3
- Wilcock, J.D., 1968b
In search of Camelot (Press report)
Computer Weekly, August 22, 1968, 2
- Wilcock, J.D., 1968c
Computer joins search for Camelot
Computer News 12(9), 1

- Wilcock, J.D., 1969a Computers and Camelot
South Cadbury, an exercise in computer
archaeology
Spectrum, British Science News 60, 7-9
Central Office of Information, H.M.S.O.,
London
- Wilcock, J.D., 1969b Computer analysis of proton magnetometer
readings from South Cadbury 1968 - A long-
distance exercise
Prospezioni Archeologiche 4, 85-93
- Wilcock, J.D., 1970a Some developments in the portrayal of
magnetic anomalies by digital incremental
plotter
Prospezioni Archeologiche 5, 55-58
- Wilcock, J.D., 1970b Prospecting at South Cadbury: an exercise
in computer archaeology
Science and Archaeology 1, 9-11
- Wilcock, J.D. and Gee, J., 1965 Tracing a Roman secondary road south from
Leek, Staffordshire
*City of Stoke-on-Trent Museum Archaeological
Society Reports* 1, 10-17

Chapter 6

MISCELLANEOUS APPLICATIONS

6. *Miscellaneous Applications*

6.1

The major archaeological topics to which the computer has been applied have been described in Chapters 1-5. There are, however, numerous minor topics which have been similarly treated, and these are described below.

6.2

General

In this section general topics are described, including reference works with broad coverage.

6.2.1

Chenhall (1965 onwards) has edited the *Newsletter of Computer Archaeology*. The scope of this publication covers all applications of the computer in archaeology.

6.2.2

Ralph (1965 onwards) has edited the *MASCA* (Museum Applied Science Center for Archaeology) *Newsletter*. This publication covers all aspects of science applied to archaeology in brief note form. There are occasional notes on computer applications in archaeology.

6.2.3

Hymes (1965) edits the book *The use of computers in anthropology*. All the papers concern the use of statistics in anthropology and some have archaeological aspects (e.g. Needham, 1965, see paragraph 1.6.7.2).

6.2.4

Chenhall (1966) gives a report of the International Symposium on Mathematical and Computational Methods in the Social Sciences (Rome, July 4-8, 1966). He reviews papers by Ankel and Gundlach (prehistoric pottery; the ARDOC system of information retrieval, in

which a 10mm grid is used to record vessels in three planes, and to store visual representations approximating to the shape of the pottery); Scollar and Linington (computer reduction of proton magnetometer readings from Etruria); Moberg (problems of trait determination of partial flint flakes or blades); Ellisseeff (Scalograms for the study of archaic Chinese bronzes); and Chenhall (the logic of models used for the processing of archaeological data).

6.2.5

Cowgill makes a general survey of international use of the computer in archaeology in two 1967 papers. The survey covers mainly North American workers, but there are a few remarks about work in Western Europe. No Russian references are quoted. The papers are reviewed by Chenhall (1968) in *Computing Reviews*.

6.2.6

Binford and Binford (1968) edit the book *New perspectives in archeology*. There are many references to the use of computers, but all of these are statistical studies (Hill, Longacre and Williams, see paragraph 1.5.12; Whallon, see paragraph 1.7.7; Cowgill, see paragraph 1.5.11).

6.2.7

Chenhall (1968) gives his assessment of the impact of computers on archaeological theory. There is a general survey of work in the U.S.A. on archaeological theory using the computer, and some remarks about the contribution which the computer has made. Three recent archaeological studies which have involved the use of computers are described. Each is placed in the theoretical framework in order to indicate what has been achieved and what remains to be accomplished. Further prospects, and factors which could possibly hinder future work are described.

D.L.Clarke (1968) publishes the major work *Analytical Archaeology*, a theoretical treatment of the subject which borrows from many disciplines, with varying degrees of relevance. A full chapter concerns computer applications in archaeology, a good review of the work published before 1968, but now inevitably dated.

Stochastic models (random walk, Monte Carlo, Markov chain) applied to archaeology are investigated. In a discussion of *time trends*, the seriation method of Brainerd and Robinson and Kendall's method of reading round the horseshoe (for linear data) are mentioned. The difficulties of non-homogeneous data, the Doppler effect and multi-linear data are also studied. For *space trends* the diffusion of cultures, network and node development, sets and logic, and trend surface analysis are discussed. An example of trend surface analysis applied to British Beaker pottery (Clark's major study) is given. The hexagonal method of sampling population regions not determined by natural boundaries is a strategy discussed for nearest-neighbour analysis.

In numerical taxonomy the objective definition of attributes, measures of similarity and the setting up of a structural hierarchy by ranking are discussed. The d-Mac pencil follower is described (see paragraph 3.3.19), and there is a mention of automatic sensors of shape.

Single-link and average-link clustering methods, with nodal clustering, factor analysis and multidimensional scaling are discussed. Various methods of presentation of the results (contoured matrices, dendrograms, 3D, 2D and 1D models) are given.

In a discussion of statistical methods, the various techniques are presented with the type of data to which they are most suited:

χ^2 (2 or more non-quantitative variables); analysis of variance (1 quantitative and 1 or more non-quantitative variables); analysis of covariance (2 or more quantitative and 1 or more non-quantitative variables); linear regression (2 quantitative variables); and multiple regression and factor analysis (3 or more quantitative variables).

In a discussion of computer methods in archaeology the work of several authors receives mention. The U.S.A. bias of previous publications is remedied in Clarke's book. The authors mentioned are: Binford and Binford (Middle Palaeolithic Mousterian assemblages, 1966; Comparison of factor analysis with Bordes' subjective classification; Isolation of five factors; See paragraph 1.3.13.); Doran and Hodson (Middle Palaeolithic Mousterian assemblages, 1966; Multidimensional scaling; The results call into question some of Bordes' classifications; See paragraph 1.7.5.); Sackett (Upper Palaeolithic artefact types, 1966; End scrapers, various angles and curves measured quantitatively; Seriation; See paragraph 1.4.13.); Binford (Late Palaeolithic projectile point types, 1963; Arrowheads are examined for the presence/absence of seven different attributes; χ^2 is used to test the correlation of attributes with particular stone types.); Clarke (British Bell Beaker pottery types, 1968; Analysis of the whole corpus of British Beaker pottery, 760 specimens, according to 39 attributes; See paragraph 4.3.2.); Deetz (Arikara Indian pottery assemblages, 1965; See paragraph 1.3.10.); Brown, Freeman and Longacre (The Carter Ranch Pueblo, 1964; See paragraphs 1.3.8 and 1.5.12.2.); Freeman (Pueblo pottery from Upper Colorado, 1962; Use of the Brainerd-Robinson technique.); Gardin (Reconstruction of an economic network of the 2nd millenium B.C., 1965; Analysis of clusters of merchants' names on clay tablets from Kültepe, Central Anatolia, revealing trade with Assur; See paragraph 2.3.4.);

Ankel and Gundlach (Taxonomy of Bronze Age vessels from Germany, 1966; Use of a grid to record the shape of a vessel in terms of three planes; Information retrieval for shapes on demand; The ARDOC system; See paragraphs 2.3.7 and 4.3.8.); Hodson, Sneath and Doran (Fibulae from Münsingen-Rain, 1966; The use of dendrograms, multidimensional scaling, single-link and average-link cluster analysis; See paragraph 1.6.6.); Fedorov-Davydov (Medieval Turkic nomads' stirrups, bits, etc., 1965); Cowgill (Teotihuacan, 1967; Mapping of an extensive area by the SYMAP program, implemented by H.T.Fisher in FORTRAN IV for the IBM 7094 machine; The use of factor analysis, cluster analysis and multiple regression; See paragraphs 1.5.8 and 1.5.11.); and Scollar and Linington (Proton magnetometer survey reduction, 1966; See paragraphs 5.2.10 and 5.2.23.). As can be seen from the arrangement above, the presentation is not well structured, but most of the important applications published up to 1967 receive some mention.

Clarke maintains that archaeology must be reorganised and remodelled, and mental habits must be changed to take advantage of the potential of the computer. As Lady Ada Augusta Lovelace, daughter of Lord Byron, and the first computer programmer (to Charles Babbage's *Analytical Engine*, Cambridge, 1843) maintained: The computer "has no pretensions whatever to originate anything. It can do whatever we know how to order it to perform".

Clarke's book *Analytical Archaeology* is reviewed by Wilcock (1970).

6.2.9

That there is an element of opposition to the use of computers among some archaeologists has already been demonstrated above (paragraphs 2.4.6 and 2.4.13). In the vanguard of this minority group is Jacquetta Hawkes. In a 1968 article she likens the advent of

technology in archaeology to the insidious undermining and domination of the servant over the master in Pinter's "The Servant". She claims that such techniques as the proton gradiometer, radio-carbon dating, statistical applications, etc. absorb money that might have been spent elsewhere, and attract a great part of the available energy, attention and intellectual love. There is also to be considered the effect of the selection which these disciplines impose on the kind of people attracted to archaeology. Jacquetta Hawkes claims that these people will not produce quality writing on humanistics, even when in middle age. Ingenious devices, exact measurements and statistical analyses are of no value in themselves. They must not appear to be dominating archaeology, or they will repel young people of "strong historical imagination" who should be recruited. Technology must not be allowed to "usurp the throne of history".

While this standpoint has some valid arguments, most of the views are extreme. It is a pity that the older archaeologists and the "moderns" should be increasingly forced to take up extreme positions in the technology v. humanistics argument, and this may well lead to a serious rift in archaeology as a whole. It is the considered opinion of this thesis that with toleration by both sides a middle course can be pursued.

6.2.10

The *Datafair 69* Conference of the British Computer Society was addressed by Wilcock (1969) on the application of the computer in archaeology. A summary of the current applications in the fields of information retrieval, reduction of site readings, and statistics (systematisation, hypotheses and models) was presented. The general scope of this work, the basis of this thesis, is also described in an entry in the Directory of Scholars Active of *Computers and the Humanities* (Wilcock, 1970).

6.2.11

The Anglo-Romanian Conference on Mathematics in the Archaeological and Historical Sciences (Mamaia, Black Sea Littoral, 16-24 September, 1970) is reported by Wilcock (1970) in *Science and Archaeology*. Another report appears in *Antiquity* (Kendall and Hodson, 1970). The official publication of the papers for this conference appeared in 1971 (Hodson, Kendall and Tautu (eds)). Several of the papers have received mention in this thesis: Doran (paragraph 1.5.15); Gelfand (1.9.2); Gower (1.3.32); Hiorns (1.6.7.5); Hodson (1.3.33); Kendall (1.6.7.3); Kruskal (1.6.8); Moberg (6.2.14); Moore (6.8.2); Orton (4.3.15); Sibson (1.3.34 and 1.4.18); Spaulding (6.2.12); Wilcock (6.2.13); and Wilkinson (1.6.7.4).

6.2.12

Spaulding (1971) gives a general summary of current techniques, including typology work by Kendall (use of the incidence matrix, the Q technique on an artefact-artefact correlation square matrix, the R technique on an attribute-attribute correlation square matrix, and factor analysis), prototype seriation work by Petrie and the developments of Robinson, Kendall, and Craytor and Johnson, numerical taxonomy work by Doran and Hodson, and attribute association work by Sackett.

6.2.13

Wilcock (1971) describes non-statistical uses of the computer in archaeology, under the headings information retrieval (the large body of specialist data; the museum collection; and the excavation record), routine reduction of instrument survey readings, graphics (d-Mac pencil follower, digital incremental plotter, advanced plotters, line-drawing display unit), the objective classification of artifacts by pattern recognition techniques, distribution maps and a software system for archaeologists.

6.2.14

Karl-Axel Moberg (1971) in his closing address to the Mamaia Conference (see paragraph 6.2.11 above) gives a review of previous conferences on the subject of mathematics in archaeology, and a perspective view of the contributions at the Mamaia Conference, mapped onto the background of archaeology.

6.2.15

A technical manual, written by Chenhall (1971), has been published by the IBM Corporation, and is intended for those anthropologists and archaeologists who know nothing about the computer, but suspect it may be able to help them in their work. The manual begins with a description of card codes. The advantages of the computer over other calculating machines are listed. It is emphasized that it is not necessary to have experience in mathematics. The problems of learning to write programs and the precise definition of data are covered, and a typical classification code book is given. The purpose of flow charts is described. Next the creation of files is covered, with descriptions of the card punch, remote terminal, paper tape punch, magnetic tape handler, document readers and graphical displays. Methods of storage covered are magnetic tapes, discs and core. The *hierarchical, keyword-in-context* (KWIC), and *tagged descriptor* methods of indexing are mentioned. On the topic of *data banks*, files of specialist information (e.g. radiocarbon dates), museum files (the Museum Computer Network), and the Arkansas Archeological Survey (an example of centralised site records) are described. The transmission of data by satellite circuit from a site in Hawaii to a computer in Palo Alto, California by Stell Newman in 1969 (see paragraph 2.3.22) is described. This would not have been possible without the design

of forms for direct keypunching, without intervention of an editor between archaeologist and punch operator to copy the data into the prescribed input format. Newman uses 3-part forms. The original is used for keypunching, the second copy is filed by coordinate and level number for use in the field, and the third copy is the permanent archive record which is filed by site, feature and artefact numbers. The *global reference code*, a system based on latitude and longitude is specified.

In information retrieval the need for thesauri to avoid synonyms is mentioned. Selective printing, listing the whole file, merging and sorting are described. On the topic of data analysis there is a very general summary of statistical techniques, viz. taxonomic distance, seriation, factor, cluster and proximity analyses (the last method deserves much greater coverage), dendrograms and chi-square. The use of the BASIC language and a remote terminal are described. There is a brief mention of terminals in the home, graphics and microfilm.

The bibliography cites most of the important contributions in the U.S.A., but very few publications by workers in Britain and Europe. The publications *Computers and the Humanities* and the *Newsletter of Computer Archaeology* are mentioned, but not *MASCA Newsletter* or *Science and Archaeology*.

6.2.16

Shorter (1971) publishes a practical guide for the historian who wishes to use the computer. In this book there is practically no overlap with archaeology, but the techniques described are common to both historical and archaeological applications. After a survey of historians who have worked with computers, the design of the codebook, and need for coding is covered. Next, card-handling and computer programming for the data processing is described. The work concludes

with descriptive statistics and remarks about correlations, with a warning about the misuse of statistics. Sample codebooks are given in appendices.

6.2.17

Gardin (1971) gives a general review of the use of the computer in archaeology since the work of Ihm in 1959 (see Ihm, 1961, paragraph 1.3.4). Gardin considers that the computer has only highlighted certain shortcomings in archaeological methodology, viz. the choice of attributes and their arrangement in a data matrix, the objectivation of the concept of resemblance, and the acceptance of numerical methods as a "cure-all". Methods of seriation are described. On the topic of information retrieval for documents, Gardin comments on the inertia and jealousy of traditional workers when faced with computerisation and its possible effects on the status of their work. He concludes with some remarks on simulation and validation studies.

6.2.18

James Doran (1972) in a discussion with Colin Renfrew, Glyn Daniel, Jacquetta Hawkes, R.J.C. Atkinson and others reviews applications of the computer in archaeology, such as information retrieval, automatic classification, seriation, simulation, models and hypotheses. Atkinson speaks of the dangers of the two methods - "one can tell lies with numbers just as easily as one can with words". Jacquetta Hawkes discounts the theory that Stonehenge was a computer (see section 6.3 below), and deplores the intrusion of scientific and technological aids into archaeology.

6.3

Astronomy and Standing Stones

There has been much discussion on the significance of standing stones in ancient religious practice, agriculture and astronomy. It was perhaps inevitable that the computer should be used to check alignments and to construct hypotheses about the uses of these stones.

6.3.1

Hawkins and White (1965) perform a computer study of the alignments at Stonehenge, on Salisbury Plain. There seems little doubt that this henge monument has some astronomical significance, for the midwinter sunset is framed accurately by the great central trilithon, and there may have been ceremonies at midwinter to "regenerate" the sun at its weakest time. Popular attention is nowadays drawn by the midsummer sunrise which is near the stone called the *Heel Stone*, although it has been shown that at the period the henge was in use it would have been further to the left; moreover, the Heel Stone was at this time partly obscured from the centre of the circle by an upright sarsen stone which is now fallen, and the rising sun was probably accurately framed between two uprights somewhat to the left of the Heel Stone. Hawkins credits the ancient priesthood with extraordinary powers of astronomical observation (within a few minutes of arc), "proving" alignments not only for the sun and moon but for certain stars. He remarks "I have demonstrated beyond reasonable doubt that the monument was deliberately, accurately, skilfully oriented to the Sun and Moon".

6.3.2

R.J.C. Atkinson (1966) in his article "Moonshine on Stonehenge" casts doubt on some of the archaeological statements made by Hawkins,

the accuracy of the plans from which alignments were taken for the computer study (Hawkins did not himself visit Stonehenge and take his own measurements), and the correctness of the statistical tests of significance. Atkinson himself has published a substantial work on Stonehenge (1956, Hamish Hamilton, London).

6.3.3

Alexander Thom (1966) has measured the alignments of many standing stones in Britain and France. He gives histograms of the predominant headings which seem to indicate a calendar split up into 16 periods. Thom was originally Professor of Engineering at Oxford University, and his work has been criticised by Jacquetta Hawkes seemingly on the grounds of his non-humanistic background, although it is accepted by many archaeologists. Thom postulates a "megalithic yard" which seems on the basis of statistical checks to be a component of many of his measurements of stone circles.

6.4

Authorship Studies

The authorship of ancient texts has been statistically checked by computer by counting word frequencies, measuring sentence lengths etc. as indicators of style. The χ^2 test is frequently used as a test of significance for checks on texts which may be of limited length.

6.4.1

Russell (1967) describes the COCOA system (a word COunt and COncordance generator for Atlas), software for generating word count lists, concordances, and KWIC indexes, and for performing statistical checks on the word counts for various books within the whole corpus of text. The input is from magnetic tape, and there is line printer output. Andrew Morton has used this software a great deal in his analysis of Greek texts (see paragraph 6.4.3 below).

6.4.2

Bodson (1967) discusses the analysis of Latin and Greek texts by computer. One of the problems is the representation of Greek letters by the Roman characters available on most line printers. It is possible to use specially-designed typewriters to punch Greek text and produce codes on paper tape for input and a *chain printer* with special characters may also be used for output. The chain printer is not restricted in character set as is the *drum printer* (in which the character set must be placed around the circumference of the print barrel); an endless chain carries the character "slugs" horizontally behind the ribbon and the print hammers impress the paper onto the slugs through the inked ribbon.

6.4.3

Morton and Levison (1968) have produced several papers on

the authorship of Greek prose. The COCOA system (see paragraph 6.4.1 above) is used to produce statistical predictions of authorship based on sentence length distribution, and the occurrence of such common Greek words as $\kappa\alpha\iota$, $\delta\epsilon$ and $\gamma\alpha\rho$. The prediction is based on the unconscious habits of writers (normally called "style") which reflect themselves in the frequency of use of common words. Thus the real differences between authors lie not in the distribution of infrequently-used words but in the words which all frequently use. It is possible on this basis to divide works into likely, unlikely and indeterminate authorship, based on the χ^2 test.

6.4.4

B.R.Schneider (1971) discusses the production of machine-readable text, a prerequisite of all text analysis by computer. Devices mentioned are the *keypunch* (punched cards), *key-to-tape typewriter* (paper tape), *key-to-magnetic tape typewriter* (magnetic tape), the *magnetic tape Selectric typewriter* (magnetic tape), the *online typewriter* (connected directly to the computer), *optical character recognition* (for both typed and handwritten characters), *printers' paper tape* (a by-product of the printers' typesetting operation), *printers' magnetic tape* (ditto), the *cathode ray tube terminal* (with keyboard) and the IMLAC CRT. The advantages and disadvantages of each device are given. Verification is necessary when cards are key-punched. The data recorded by key-to-magnetic tape requires no further conversion before input to the computer. The magnetic tape Selectric typewriter has editing facilities also, but costs \$10,000. Key-to-paper tape produces a paper tape for input to the computer. The online typewriter is wasteful of machine time. Optical character recognition has real

advantages, whether of printed text or text retyped with an OCRB golfball, Prestige Elite 72, Perry Font 157, or Courier 72. Printers' paper tapes or magnetic tapes are a by-product of the compositor, thus removing the data preparation stage. Magnetic tape is used for photocomposition, paper tape for the Linotype machine. The alphanumeric CRT terminal or a line-drawing display unit (e.g. the IMLAC CRT) may be used to input codes direct to the computer. They are not as wasteful as the online typewriter in computer time, since a whole frame of characters can be built up in the display buffer before transmission. The displayed characters (for visual effect only) are generated by hardware or, in the case of the line-drawing display unit, by software also if desired. CRT terminals are in general the best medium for input of text, with their visual images, comprehensive editing functions before transmission, and complete interaction between man and machine.

6.5

Decipherment of Ancient Languages

In this application the computer treats the ancient language as a code to be cracked. There may be a requirement for the representation in code of a large number of symbols, particularly in hieroglyphic languages.

6.5.1

Mackay (1965) describes the type-found of the *Phaistos disc*, a well-known artefact from Crete. By statistical analysis he deduces that the number of symbols was about 55 in all, and that the unknown language is possibly of the consonant-and-vowel type e.g. Linear B. There is insufficient text to form a basis for actual decipherment. Mackay discusses the method of writing, which appears to have been by impressing pottery dies into the clay disc.

6.5.2

Morgan and Reich (1966,1967) report Assembler Language software for the IBM 360/65 machine for the analysis of the Minoan hieroglyphic script. All Minoan hieroglyphic inscriptions (seal stones, sealings and graffiti) are to be examined, punched onto cards and transcribed to magnetic tape. The symbols are to be listed and sorted for correlations with provenances, materials and colours of the seal stones. Formulae involving two or more symbols are to be listed and correlated.

6.5.3

Clauson and Chadwick (1969) describe the methods of Finnish workers who may have partially deciphered the Indus script, plausibly arguing that it is a form of proto-Dravidian. The supposition rests on the identification of signs for plural, genitive, dative, man and woman based on the homophones in Dravidian for bamboo pole, ship,

arrow, comb, etc. A computer is used to perform the statistical counting of the symbols, which were all given a 3-digit code. The number of occurrences of symbols, first and last symbols in words, and pairs of symbols are counted, and key-sign-in-context lists output, but no other analysis by computer has so far been attempted. This counting could equally well have been performed by hand.

6.5.4

Gardin (1969) reports FORTRAN IV and V software for the UNIVAC 1108 and IBM 360/44 machines for the analysis of written documents, based on syntax, semantics, factor analysis and graphs.

6.5.5

Thomson (1972) reports COBOL F software for the IBM 360 system for the input and analysis of 30,000 Sumerian cuneiform economic tablets. KWIC listings and classified records are produced.

6.6

Egyptology

The applications of the computer in Egyptology have concerned aspects of the great monuments erected by the Egyptians.

6.6.1

The University Museum of Philadelphia (1966), and the U.S. Government through the Smithsonian Institution have sponsored a project to recreate on paper the Temple of Akhenaten at Karnak. This temple was erected by the heretic Pharaoh Akhenaten in honour of the Aten and was largely dedicated to his consort Queen Nefertiti. Akhenaten also set up a new capital at Tell-el-Amarna. After his death the priests of Amun reasserted themselves (the Pharaoh Tutankhaten changed his name to Tutankhamun), the temple of Akhenaten was razed to the ground and the massive blocks of masonry were incorporated in the foundations of pylons and other buildings. Many of the blocks carry carvings, and the project is to reassemble these like a gigantic jigsaw puzzle. The divergent rays of the Aten are particularly helpful in the matching of adjacent blocks. Many of the blocks are stacked up at Karnak, but a large number have been dispersed to museums throughout the world. All stones are to be individually photographed and the data fed into a computer. The problem is then one of information retrieval and matching, and the expectation is that the blocks can eventually be grouped in their original order so that the inscriptions and polychrome relief can be recovered.

Most of the publications describing the project are by R.W.Smith (1966,1968,1970). He states "the data on each stone will be fed into a computer with the expectation that they can eventually be grouped into their original order. This is a very simple process for the

computer if the individual stones are properly processed". About 35,000 blocks (about one fifth of the total stones) are to be processed. The blocks are sorted under major decoration types. Some successful matches are illustrated.

6.6.2

Alvarez (1969) describes the use of cosmic ray *spark chambers* to detect the presence or absence of hidden cavities in the pyramid of Chephren at Giza. The spark chambers were located in the Belzoni chamber near the bottom centre of the pyramid, and were used to detect cosmic ray muons penetrating the pyramid. Near the bottom of the pyramid, more muons get through the sides than through the thicker masonry forming the peak of the pyramid. If a significant void existed in the pyramid, more muons than predicted would arrive from that direction. The effective volume scanned was up to 35° from the vertical in all orientations. Computer analysis of the data failed to reveal any hidden cavities in Chephren's pyramid. Similar observations are to be carried out on the lower zones of the pyramid.

6.7 Languages

The programs for this thesis have been written in the ALGOL 60 and BASIC languages. These languages are formally defined in the publications listed below.

6.7.1

The syntax and semantics of ALGOL 60 are described by Backus (1959) in terms of a specially-developed algebraic notation which has since become known as the *Backus-Naur form*. The complete definition of ALGOL 60 was presented to the 1963 IFIP Conference (edited by Naur). ALGOL 60 is a language suitable for the expression of a large class of numerical processes in a form sufficiently concise for direct automatic translation into the languages of programmed automatic computers. The introduction to the report contains an account of the preparatory work leading up to the final conference. The Backus-Naur form is used to define the *reference language*, and the *publication language*. ALGOL 60 does not cover input/output instructions in any detail. It has to be implemented on different computers according to their hardware configurations, with the result that different implementations of ALGOL 60 are quite dissimilar in their input/output conventions, being similar only in the arithmetic, logical and control features. It is possible to extend ALGOL 60 greatly by adding sophisticated input/output routines, an example of which is ICL EDGAR for control of a display unit.

6.7.2

The BASIC language is formally defined by Lee (1972). He gives details of the operation of an abstract definition machine, containing an *analyser*, *translator* and *interpreter*. The analyser verifies the validity of the text and generates a parsed text. The

translator takes the parsed text and transforms it into abstract text using default conditions, which are given in full. The syntax is specified in Backus-Naur form. BASIC is a language designed for remote terminal use. All statements are numbered, and each is checked by the analyser as it is input. Programs and data may be recorded permanently on disc by the command SAVE, and retrieved at any time for running. A standard program may be retrieved, and the program and/or data modified by typing in statements before running, but the standard program remains unchanged on the disc unless the command SAVE is used. This operating system provides the maximum of man-machine interaction from a remote teletype terminal, but is costly in terms of storage and CPU involvement.

6.8 *Mosaics*

Mosaics have been used to pave floors of buildings in antiquity. They were first developed by the Greeks and later elaborated by the Romans, although it appears from the analyses described below that the Greek standards of measurement were retained by the mosaic workers in Roman times.

6.8.1

The first, surprisingly early, mathematical analysis of mosaics is by Sylvester (1867). This is a mathematical treatment of tessellation, based on sign-successions in matrices. A *self-reciprocal matrix* is defined as a square matrix in which each element is proportional to its first minor. A strictly *orthogonal* matrix has a *determinant* equal to unity, but any self-reciprocal matrix may be termed orthogonal if the concept of strictness is withdrawn. If each element is the inverse of its first minor, the matrix is termed an *inverse orthogonal matrix*. This is found to be connected with the calculus of sign-progressions suggested by the form of Newton's rule, leading to a theory of tessellation "highly curious in itself". Although the mathematics of this paper is interesting, it seems highly unlikely that the ancient mosaic workers would have had such thoughts in their heads when they were constructing the mosaics.

6.8.2

All the recent work on the mathematical analysis of mosaics has been by one worker, R.E.M. Moore (1966, 1968, 1969, 1971). Moore has made several hundreds of thousands of measurements on ancient mosaics and the statistical analysis reveals *modal* (i.e. predominant) lengths of 2.4, 3.6, 6.0, 9.6, 15.6, (21.6), 25.1, 40.7, 65.8 and 106.5 cm.

Apart from the bracketed measurement these form a Fibonacci series, which is at least some indication that there is some system involved. The above values are called "*mosaic units*" by Moore, and are possibly based on measurements on a Standard (Greek) measure. Packing experiments were carried out, using random numbers and dice to determine the choice of ancient stones and their orientation, and a specially-designed packing machine funded by IBM and the BBC. The machine has a hopper, conveyor belt, reference edge and rubber strips to define the channels. The results confirm the existence of "*mosaic units*" and on these Moore makes deductions on the calibrations of ancient mosaicists' rulers. There is much evidence against the use of specially-cut stones and for the use of "standard" stones of modular dimensions.

Pollen Analysis

Pollen analysis is used by archaeologists to determine ancient climates. Suitable peat bogs preserve pollen grains of the species of plants existing at the time the peat was laid down, and a continuous sequence over hundreds of thousands of years is sometimes obtained, in effect a "time capsule". Pollen grains have an outer casing which is virtually indestructable, even in peat acids. The Neolithic farming revolution is indicated in the pollen spectra by a reduction in tree pollens (the "Ulmus decline") and an increase in agricultural weed pollens, indicating forest clearance. The actual numbers of pollen grains of each species are counted under a microscope. The variations in predominance of each species are presented in a final publication as graphs or histograms of percentages of the pollen sums for each group of species (e.g. trees, shrubs, herb pollen, spores) and of the total pollen count. This is a routine statistical and graphical exercise well-suited as an application for computers. A computer program for pollen analysis has been written for this thesis in collaboration with Miss K. Simpkins, a pollen analyst at the University of Keele (see paragraph 6.12).

6.9.1

Professor G.W.Dimbleby (1969) of the Institute of Archaeology, University of London, presents a general survey of pollen analysis methods.

6.9.2

R.H.Squires of the University of Durham has developed a computer program for pollen analysis (Squires,1970; Hill and Squires, 1970; Squires and Holder,1970). This is written in FORTRAN IV for

the IBM 1130 machine and requires a line printer and graph plotter. Squires claims that this is the first attempt to present pollen data by an automatic process (however, ARCJW015, the pollen analysis program for this thesis was developed concurrently in 1970). Squires' publications are aimed at the computer scientist and his descriptions of the programs would not be intelligible to the average archaeologist. The output routine has been developed from a program by C.W.A.Browitt of the University of Durham, which plots seismic traces adjacent to one another and uses standard plotting subroutines supplied by IBM. The output is in the form of several parallel graphs with the depth of the sample vertical and the percentages of the pollen sum horizontal. Histograms are not used (although this seems to be the preferred presentation for archaeologists), as is the case in ARCJW015 (see paragraph 6.12). Percentages of the pollen sum are tabulated on the line printer with abbreviated names of the species. The input is by punched card, and the percentages are calculated of a *specified* total (this is not calculated automatically as in ARCJW015), and suitable scales must be specified for the x and y axes of the plotter (again, these are not determined automatically as in ARCJW015). The number of taxa must also be specified (not automatic) and only four letters are allowed for a taxon name. There are limits of 100 taxa and 100 horizons (depth samples). Squires suggests the setting-up of a data bank for pollen data.

6.10

Provenance

A basic requirement of the archaeologist is to determine the provenance of an object; in the case of pottery, metallic objects etc. this may be done by determining the amounts of various trace elements in the material, thus reflecting the origin of the raw materials. The similarity of various objects may then be calculated as regards constituents and compared to samples of raw materials from various localities. The computer has been used in performing the various statistical calculations necessary and in producing graphical outputs.

6.10.1

Newton and Renfrew (1969) use the computer to apply *element analysis* to data on the faience beads found in Britain. They use a statistical technique which retains the variables in their original form and uses correlations to indicate the coefficients in a regression equation. In particular, the method indicates which variables produce the best discrimination. 75% success is indicated by the computer for discrimination between the areas of origin of the material on the basis of the amounts of tin, magnesium and aluminium. The statistical separation of faience beads from Scotland, Wessex and Egypt makes the theory of importation of faience beads from the Mediterranean less credible. The authors consider that it is likely that the British beads were made in Britain. It seems that the appearance of faience beads coincides with the development of bronze technology, and perhaps the newly-developed skills with alloys led to the invention of faience.

6.10.2

Sieveking, Craddock, Hughes, Bush and Ferguson (1970) describe the characterisation of British flint mine products. Data for the

concentrations of aluminium, magnesium, potassium and iron in flint samples from six localities are studied by plotting the concentration of one element against another. Some localities occupy distinct fields in this representation. Factor analysis using the method of principal components, and discriminant analysis is carried out on the CDC 6600 machine using modified versions of the programs given in the IBM System/360 manual (Scientific subroutine package 360A-CM-03X Version 2, Programmers Manual, IBM, New York, 1967). It is found that flints can be allocated to sites of origin on the basis of this analysis with up to 95% accuracy.

6.10.3

Beck reports FORTRAN IV software for the IBM 360/30 machine for the determination of the origin of Greek amber artefacts by computer-classification of infrared spectra (Beck, 1970; Beck, Adams, Southard and Fellows, 1971). The amber artefacts are submitted to physical analysis by a number of multiparameter methods, principally *infrared spectroscopy* and *nuclear magnetic resonance*, followed by computer classification into geographical groups. The computer storage and retrieval of infrared spectra is not new, but previous use has mainly been in organic chemistry. The readings are divided by the absorbance at $8.0\text{ }\mu\text{m}$ to cancel out fluctuations caused by different thicknesses or concentrations. Treating very short portions of the spectral curve as straight lines, the difference between neighbouring absorbances is proportional to the slope of the absorption curve. Four aggregate slopes are taken for the spectral bands $8.1 - 8.3\text{ }\mu\text{m}$, $8.3 - 8.5\text{ }\mu\text{m}$, $8.5 - 8.7\text{ }\mu\text{m}$ and $8.7 - 8.9\text{ }\mu\text{m}$. Each slope is expressed as 0, + or - . Baltic amber gives the sequence 0,0,-,+ and oxidised Baltic amber gives -,-,-,+.

The computer is used to generate all possible slope sequences, and

an attempt is made to allocate these to real samples. Finally a classification scheme is evolved based on the slopes of the absorption curve in specific bands, usually in the 0,+,- form but occasionally based on absolute values of slope.

6.10.4

Fields, Milsted, Henrickson and Ramette (1971) use trace impurity patterns in copper ores and artefacts to deduce provenance. The artefacts and ores are submitted to physical analysis by optical emission spectroscopy, neutron activation analysis, atomic absorption and spark source mass spectroscopy. A statistical approach is employed, concentrations being expressed on a "unit scale" (sub-logarithmic), based on ranges from 100%-32% down to 1.0×10^{-8} - 3.2×10^{-9} . A computer program has been developed which can retrieve the impurity concentration of an ore, smelt or artefact. A library of impurity "fingerprints" for ores and smelts is being built up, so that the computer can compare the "fingerprint" of an artefact to all the data and determine the most likely geographic origin of the copper used in the artefact. Comparisons with other artefacts are also made.

6.11

Radiocarbon Dating

The radiocarbon dating method has long been used to determine the absolute date of a site or artefact. The method retains its importance, although the *Suess (fossil fuel) effect*, the *atomic bomb effect*, and latterly the variation in cosmic ray intensity as revealed by the *Bristlecone Pine studies* have shown that it is not as easy a method as was first thought. The computer has recently been used for information retrieval purposes in connection with radiocarbon dates, and in the compilation of chronologies.

6.11.1

Ferguson (1967) reports the finding of an additional specimen of the Bristlecone Pine (*Pinus aristata*) in the White Mountains of eastern-central California which has extended the tree-ring chronology from 6600 to well over 7100 B.P. The computer is used to compile a White Mountains chronology.

6.11.2

Taylor, Berger and Dimsdale (1968) describe electronic data processing for radiocarbon dates. The need for a more flexible index for radiocarbon dates with shorter retrieval times has led to a suggestion for the development of a computer or punched-card data retrieval system. A suggested format for coding radiocarbon data on punched cards and two techniques for the rapid retrieval of desired information are discussed.

6.11.3

Ferguson (1970) describes the use of the computer to filter the fluctuations in the Bristlecone Pine calibration curve and to plot a 7484-year chronology of the White Mountains. Measurements on tree-rings are processed by standard programs.

6.11.4

Suess (1970) mentions in discussion that he is using the computer to analyse the power spectrum of the radiocarbon calibration curve, to eliminate high frequencies and to determine the highest frequencies which need be considered. Finally, the aim is to instruct the computer to draw an unbiased calibration curve.

ARCJWO15 Pollen analysis (ALGOL 60)

This program performs routine statistical calculations on pollen data and outputs the results in histogram form. The data is read from paper tape. First the number of samples taken is specified, then the site identifier is read using procedure INSTRING. The period of the samples, whether late-glacial or post-glacial, is specified. In late-glacial diagrams the pollen sum is usually based on the totals of tree, shrub and herb pollen *excluding* aquatic plant pollen and spores; in post-glacial diagrams the pollen sum is based solely on the tree pollen totals. This procedure has some illogicality as it sometimes leads to percentages of pollen sum in excess of 100.0, a point which has been pointed out to the originator of the data, Miss K.Simpkins of the University of Keele, and this may lead to a modification of technique in this respect.

The depths of the samples in cm are specified next. The names of the species, followed by their pollen counts for each of the sampling depths are read. Each name is terminated by * and may be a maximum of 30 characters in length. If the first character of a name is * this is taken as a signal that the data is complete. A master magnetic tape containing the full species list originally set up by ARCJWO06 (see paragraph 2.6.2) is initialised. A group identifier is cleared for each species read, a card image is read from the master tape and the species is checked for presence in the input list. Procedure SEARCH is used to do this check. If the species on magnetic tape is present in the input list, the message PRESENT is output to the line printer and the pollen group

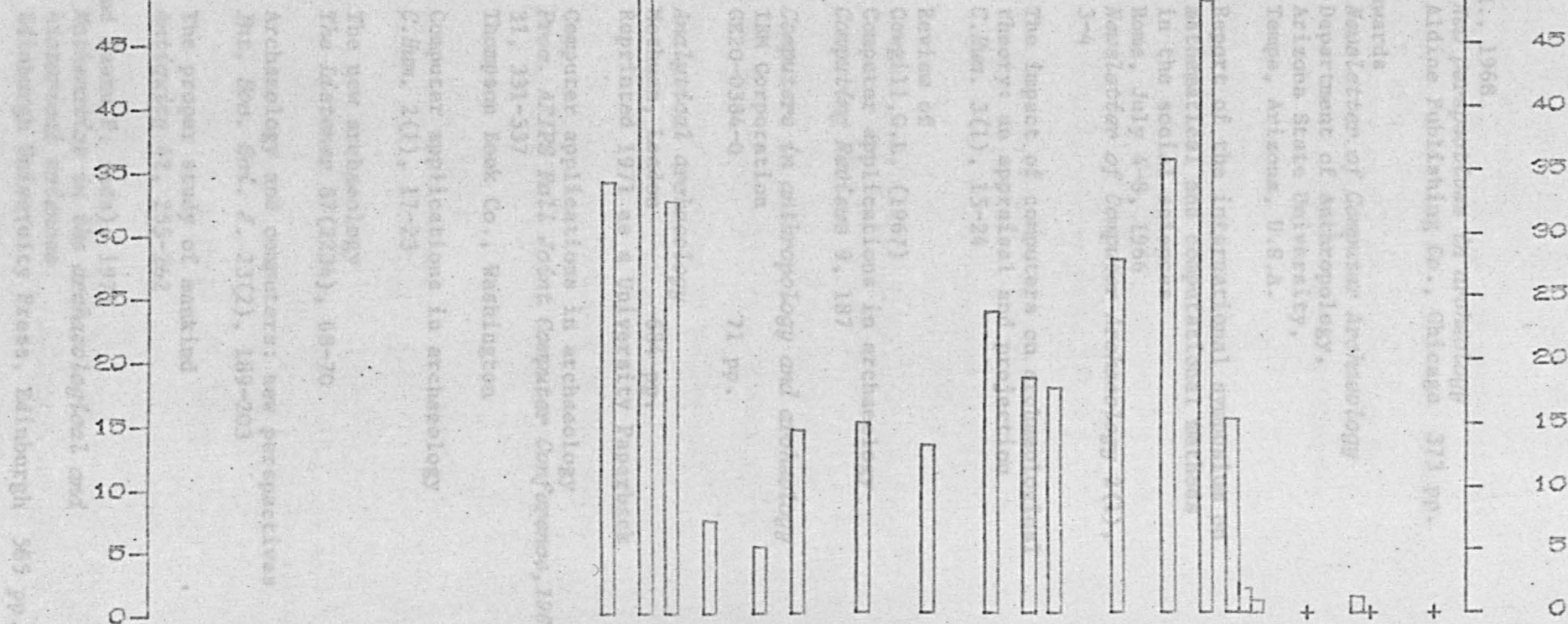
as defined by the master tape is entered in the group identifier for the species. As the input list is scanned those species for which the pollen group has already been determined are ignored. To check that the master tape is not corrupt, a check digit is calculated from the second letter of the pollen group name, as follows:

<i>Pollen group name</i>	<i>Second letter</i>	<i>Code value</i>	<i>Check Digit</i>
SHRUB	H	40	0
HERB POLLEN	E	37	4
SPORES	P	48	48-47 = 1
AQUATIC POLLEN	Q	49	49-47 = 2
TREE	R	50	50-47 = 3

The second letter rather than the first was chosen since this gives a unique code for every pollen group and lends itself best to decoding. The check digit is itself checked to be in the range 0-4 and if it is outside this range the message POLLEN GROUP UNDEFINED is output, and the species omitted from the analysis. The pollen counts for each depth are added to the corresponding pollen group sums, then the program returns to read a new card image from magnetic tape. When all the master tape has been read, a check is made that all the species have been recognised (N.B. a species name may be abbreviated if there is no chance of ambiguity). If not, the input list is scanned for the unidentified species and the message SPECIES NOT RECOGNISED is output for these, with the names as input. The program then returns to the input phase.

If all input species have been recognised, the correct pollen sum (depending on whether the analysis is for late-glacial or post-glacial times) is calculated for each depth. The x-increment for the diagram is calculated automatically, and the site name, period and depth scale are plotted, together with markers for the sampling

depths. The pollen groups are plotted in the order TREE, SHRUB, HERB POLLEN, AQUATIC POLLEN, SPORES and appropriate messages are output at the point where species are excluded from the pollen sum i.e. after the trees for post-glacial studies and after the herb pollen for late-glacial studies. The pollen counts for each depth are printed as percentages of the pollen sums for the group and the total pollen sums. The program uses procedures INSBL, INCDI (magnetic tape handling), SEARCH (identification of species in master list) and HISTOGRAM (plots a pollen diagram if the percentages are not negligible). A typical pollen diagram is shown in Figure 6.1.



Typical HISTOGRAM output for pollen analysis

Miscellaneous Topics Bibliography

General

- Binford, S.R. and Binford, L.R., 1968
New perspectives in archeology
Aldine Publishing Co., Chicago 373 pp.
- Chenhall, R.G. (ed.), 1965 onwards
Newsletter of Computer Archaeology
Department of Anthropology,
Arizona State University,
Tempe, Arizona, U.S.A.
- Chenhall, R.G., 1966
Report of the international symposium on
mathematical and computational methods
in the social sciences
Rome, July 4-8, 1966
Newsletter of Computer Archaeology 2(1),
3-4
- Chenhall, R.G., 1968a
The impact of computers on archaeological
theory: an appraisal and projection
C.Hum. 3(1), 15-24
- Chenhall, R.G., 1968b
Review of
Cowgill, G.L. (1967)
Computer applications in archaeology
Computing Reviews 9, 187
- Chenhall, R.G., 1971
Computers in anthropology and archeology
IBM Corporation
GE20-0384-0 71 pp.
- Clarke, D.L., 1968
Analytical archaeology
Methuen, London 684 pp.
Reprinted 1971 as a University Paperback
- Cowgill, G.L., 1967a
Computer applications in archaeology
Proc. AFIPS Fall Joint Computer Conference, 1967
31, 331-337
Thompson Book Co., Washington
- Cowgill, G.L., 1967b
Computer applications in archaeology
C.Hum. 2(1), 17-23
- Doran, J.E. et al., 1972
The new archaeology
The Listener 87(2234), 68-70
- Gardin, J.C., 1971
Archaeology and computers: new perspectives
Int. Soc. Sci. J. 23(2), 189-203
- Hawkes, J., 1968
The proper study of mankind
Antiquity 42, 255-262
- Hodson, F.R., Kendall, D.G. and Tautu, P. (eds), 1971
*Mathematics in the archaeological and
historical sciences*
Edinburgh University Press, Edinburgh 565 pp.

Hymes,D. (ed.), 1965

The use of computers in anthropology
Mouton & Co., The Hague, London and Paris
558 pp.

Moberg,C.A., 1971

Closing address:
Archaeological context and mathematical
methods
in Hodson,F.R., Kendall,D.G. and Tautu,P.
(eds), 1971
*Mathematics in the archaeological and
historical sciences* 551-562
Edinburgh University Press, Edinburgh

Ralph,E.K. (Associate Director), 1965 onwards

MASCA Newsletter
Museum Applied Science Center for
Archaeology
The University Museum,
University of Pennsylvania,
33rd and Spruce Streets,
Philadelphia, Pennsylvania 19104

Shorter,E., 1971

*The historian and the computer
A practical guide*
Prentice-Hall Inc., Englewood Cliffs,
New Jersey 149 pp.

Spaulding,A.C., 1971

Some elements of quantitative archaeology
in Hodson,F.R., Kendall,D.G. and Tautu,P.
(eds), 1971
*Mathematics in the archaeological and
historical sciences* 3-16
Edinburgh University Press, Edinburgh

Wilcock,J.D., 1969

Computer applications in archaeology
Datafair 69
Abstracts of BCS Symposium Papers

Wilcock,J.D., 1970a

S110. Computer applications in archaeology
C.Hum. 4(5)
in *Directory of Scholars Active 1970*, 354

Wilcock,J.D., 1970b

Review of
Clarke,D.L. (1968)
Analytical archaeology
Science and Archaeology 1, 29

Wilcock,J.D., 1970c

Our man in Mamaia
(Report of the Anglo-Romanian Conference on
Mathematics in the Archaeological and Historical
Sciences, Black Sea Littoral, 16-24 September
1970)
Science and Archaeology 2/3, 12-16

Wilcock,J.D., 1971

Non-statistical applications of the computer
in archaeology
in Hodson,F.R., Kendall,D.G. and Tautu,P. (eds),
1971
*Mathematics in the archaeological and historical
sciences* 470-481
Edinburgh University Press, Edinburgh

Astronomy and Standing Stones

- Atkinson, R.J.C., 1966 Moonshine on Stonehenge
 Antiquity 40, 212-216
- Hawkins, G.W. and White, J.B., 1965
 Stonehenge decoded
 London
- Thom, A., 1966 Time-keeping with standing stones
 New Scientist 32(527), 719-721

Authorship testing

- Bodson, A., 1967 Latin and Greek by computer
 Computer Weekly 21/28 December 1967, 6-7
- Morton, A.Q. and Levison, M. Some indicators of authorship in Greek prose
 in
 The computer and literary style
- Morton, A.Q. and Levison, M., 1968
 Literary uses of the computer
 New Scientist 15 August 1968, 340-342
- Russell, D.B., 1967 COCOA Manual (a word Count and
 Concordance generator for Atlas)
 Science Research Council, Atlas Computer
 Laboratory, Chilton
- Schneider, B.R., 1971 The production of machine-readable text:
 Some of the variables
 C.Hum. 6(1), 39-47

Decipherment of Ancient Languages

- Clauson, G. and Chadwick, J., 1969
 The Indus script deciphered?
 Antiquity 43(171), 200-207
- Gardin, J.C., 1969 S93. Méthodes mathématiques et linguistiques
 en archéologie
 C.Hum.
 in Directory of Scholars Active 1968/69(2),
 141
- Mackay, A., 1965 On the type-fount of the Phaistos disc
 Statistical Methods in Linguistics 4, 15-25
 Skriptor Fack, Stockholm
- Morgan, R.S. and Reich, J.J., 1966
 S16. Analysis of the Minoan hieroglyphic
 script
 C.Hum.
 in Directory of Scholars Active 1966/67, 233

- Morgan, R.S. and Reich, J.J., 1967
 Sl6. Analysis of the Minoan hieroglyphic script
C.Hum.
 in Directory of Scholars Active 1967/68(1), 89
- Thomson, G.A., 1972
 TP3. Ancient history and archaeology
 (analysis of 30,000 Sumerian cuneiform tablets)
C.Hum. 6(3)
 in Humanities Programs Available 185-186

Egyptology

- Alvarez, L.W., 1969
 Cosmic rays and the pyramids
MASCA Newsletter 5(1), 3
- Smith, R.W., 1968
 Cameras and computers study ruins of ancient Egyptian temple
New York Times 14 October 1968, 6
- Smith, R.W., 1970
 Computer helps scholars recreate an Egyptian temple
National Geographic Magazine 138(5), 634-655
- University Museum of Philadelphia, and the U.S. Govt. through the Smithsonian Institution, 1966
 Akhenaten Temple project: Karnak
MASCA Newsletter 2(1), 2

Languages (for computer use)

- Backus, J.W., 1959
 The syntax and semantics of the proposed international algebraic language of the Zurich ACM-GAMM Conference
Proc. International Conference on Information Processing 1959 125-132
- Lee, J.A.N., 1972
 The formal definition of the BASIC language
Computer Journal 15(1), 37-41
- Naur, P. (ed.), 1963
 Revised report on the algorithmic language ALGOL 60
Computer Journal January 1963, 349-367

Mosaics

- Moore, R.E.M., 1966
 Pattern formation in aggregations of entities of varied size and shapes as seen in mosaics
Nature 209(5019), 128-132
- Moore, R.E.M., 1968
 Test of the alignment hypothesis as the reason for the mosaic unit phenomenon
Nature 217, 482-484
- Moore, R.E.M., 1969
 Mosaic units
Archaeometry 11, 145-158

- Moore, R.E.M., 1971
A relationship observed between mosaic units and the sizes of Roman mosaic stones in Hodson, F.R., Kendall, D.G. and Tautu, P. (eds), 1971
Mathematics in the archaeological and historical sciences 445-452
Edinburgh University Press, Edinburgh
- Sylvester, J.J., 1867
Thoughts on inverse orthogonal matrices, simultaneous sign successions, and tessellated pavements in two or more colours, with applications to Newton's Rule, ornamental tile-work, and the theory of numbers
Phil. Mag. Series 4, 34, 461-475
- Pollen Analysis*
- Dimbleby, G.W., 1969
Pollen analysis in Brothwell, D. and Higgs, E. (eds), 1969
Science in Archaeology
Second edition 167-177
Thames and Hudson, London
- Hill, A.G. and Squires, R.H., 1970
Graphs drawn by computer
AREA 4, 48-52
Institute of British Geographers
- Squires, R.H., 1970
A computer program for the presentation of pollen data
Dept. of Geography, University of Durham
Occasional papers series, 11 53 pp.
- Squires, R.H. and Holder, A.P., 1970
The use of computers in the presentation of pollen data
New Phytol. 69, 875-883
- Provenance of objects*
- Beck, C.W., 1970
S107. The archaeological provenience of amber artifacts
C.Hum. 4(5)
in *Directory of Scholars Active 1970*, 353
- Beck, C.W., Adams, A.B., Southard, G.C. and Fellows, D., 1971
Determination of the origin of Greek amber artifacts by computer-classification of infrared spectra
in Brill, R.H. (ed.), 1971
Science and Archaeology 235-240
MIT Press, Cambridge, Massachusetts and London 288 pp.
- Fields, P.R., Milsted, J., Henrickson, E. and Ramette, R., 1971
Trace impurity patterns in copper ores and artifacts
in Brill, R.H. (ed.), 1971
Science and Archaeology 131-143
MIT Press, Cambridge, Massachusetts and London 288 pp.

- Newton, R.G. and Renfrew, C., 1969
Origins of British faience beads
MASCA Newsletter 5(1), 3
- Sieveking, G. de G., Craddock, P.T., Hughes, M.J., Bush, P. and Ferguson J., 1970
Characterization of prehistoric flint mine products
Nature 228, 251-254
- Radiocarbon Dating and Dendrochronology*
- Ferguson, C.W., 1967 Dendrochronology
MASCA Newsletter 3(2), 2
- Ferguson, C.W., 1970 Dendrochronology of bristlecone pine, *Pinus aristata*.
Establishment of a 7484-year chronology in the White Mountains of eastern-central California, U.S.A.
in Olsson, I.U. (ed.), 1970
Radiocarbon variations and absolute chronology (Proc. Nobel Symposium 12)
John Wiley & Sons, Inc.
- Suess, H.E., 1970 Discussion
in Olsson, I.U. (ed.), 1970
Radiocarbon variations and absolute chronology (Proc. Nobel Symposium 12) 332
John Wiley & Sons, Inc.
- Taylor, R.E., Berger, R. and Dimsdale, B., 1968
Electronic data processing for radiocarbon dates
American Antiquity

Chapter 7

The PLUTARCH System

The P L U T A R C H System

(Program Library Useful To ARChaeologists)

While the techniques described in Chapters 1-5 are powerful in themselves, they become even more powerful when combined by interactive graphics and overlays. Not only can the operator obtain true man-machine interaction, but the individual techniques can communicate via global storage. This is the concept behind the PLUTARCH Control System, which effectively creates an archaeological command language.

In the following summary, which includes operating instructions and information, a description of the programming method, details of linkage between modules, and data checking procedures with error recovery, the modules are named but not explained in very great detail. A full explanation of the purpose of the modules may be found in Chapters 3-5.

The system as described contains graphics, instrument survey reduction and simple statistical procedures. The format is specifically designed, however, to accommodate any number of modules and the addition of advanced statistics and information retrieval modules is planned. The paper tape input module is a temporary measure which may easily be replaced in future by information retrieval input.

PLUTARCH System Index

7.1 Operating Instruction and Information

7.1.1 Diagrams Segment

7.1.1.1 Keyboard Options

- A Vertical A Proportions Frame
- C Explicit Coordinates
- D Delete Display
- F Rectangular Frame
- H Horizontal A Proportions Frame
- M Multidimensional Scaling Plot
- N Normal Line Generation
- O Circle Generation
- P Phenon Dendrogram Plot
- R Reset Frame Holds
- S Sketch
- T Tracking Cross ON
- U Tracking Cross OFF
- X Exit to Global

7.1.1.2 Function Key Options

- 1 Specify Point
- 2 Previously-defined Point
- 3 Draw Line
- 4 Draw Line to previously-defined point
- 5 Delete
- 6 Text
- 7 3D Rotation and Perspective

Function Key Options

- 1 Light Button Mode
- 2 Read Tape
- 3 Sketch
- 4 Reverse Perspective Marker
- 5 Delete
- 6 Text
- 7 Reset Frame Holds
- 8 Exit to Diagrams Segment

- 8 Read Pot Profile

7.1.2 Histogram and Piechart Segment

- 7.1.2.1 Automatic Histogram
- 7.1.2.2 Automatic Piechart
- 7.1.2.3 Histogram
- 7.1.2.4 Piechart
- 7.1.2.5 Pollen Diagram

7.1.3 Maps Segment

- 7.1.3.1 Automatic entry to Inputs Segment for input of map from magnetic tape
- 7.1.3.2 Map outline and distribution symbols

7.1.4 Instrument Plot Segment

7.1.5 Permutation Segment

7.1.6 Legends Segment

7.1.7 Inputs Segment

- 7.1.7.1 Pollen Data
- 7.1.7.2 Automatic Input
- 7.1.7.3 Pot Profile Input
- 7.1.7.4 Closed Figure Input
- 7.1.7.5 Instrument Plot Input
- 7.1.7.6 Maps Input

7.2 Description of Method

- 7.2.1 Global Procedures
- 7.2.2 Primary Tree
- 7.2.3 Diagrams Segment
- 7.2.4 Histogram and Piechart Segment
- 7.2.5 Maps Segment
- 7.2.6 Instrument Plot Segment
- 7.2.7 Permutation Segment
- 7.2.8 Legends Segment
- 7.2.9 Inputs Segment

7.3 Segment Flowcharts

7.4 Coding

7.5 Data Checking Procedures and Error Recovery

- 7.5.1 Global Error Recovery
- 7.5.2 Diagrams Segment Error Recovery
- 7.5.3 Legends Segment Error Recovery
- 7.5.4 Inputs Segment Error Recovery

7. *System Specification*

PLUTARCH

Program Library Useful To ARChaeologists

The system is designed as a flexible format into which graphics facilities for archaeologists can be inserted as necessary. It is capable of future expansion to include information retrieval and statistical routines, which will have full communication with each other and with the graphics facilities.

7.1 *Operating Instructions and Information*

The following operating instructions refer to the PLUTARCH system as available in April 1973 on the Keele University ICL 4130 computer, fitted with card reader, paper tape reader and punch, line printer, digital incremental plotter, 4 magnetic tape handlers, 2 disc handlers, operating teletype and line-drawing display unit.

1. Load disc 9

2. Message DIN,PLUTARCH,9.

Segments loaded from disc.

3. Message PLUTARCH.

Calls PLUTARCH global segment and routines DRO (general), DR30 (magnetic tape handling) and DISMAN (display manipulation).

Global options displayed on screen, as follows:

*** PLUTARCH SYSTEM ***

DIAGRAMS

HISTPIE

MAPS

INSTPLOT

PERM

INPUTS

LEGENDS

See also Figure 7.1. This menu can easily be extended to include other options as they become available, in particular statistics and information retrieval options.

4. Select required option by pointing the activated lightpen at the corresponding item in the menu. The command is acknowledged by the appearance of a bright-up border around the item selected.

Segments which may be entered are as follows:

- Diagrams
- Histograms and Piecharts
- Maps
- Instrument Survey Plot
- Permutation of figures
- Input of data
- Legends

The options available in each segment are described below.

7.1.1 Diagrams Segment

The loading of this segment is indicated by the presence of the tracking cross. Options are selected by typing a single letter on the control keyboard or by pressing a function switch. Each option may be called several times.

7.1.1.1 Keyboard options

- A Rectangular frame, larger dimension vertical, in A proportions
- C Specify Coordinates explicitly. A graduated border is temporarily displayed, allowing the operator to judge the required position visually.

Teletype message: POINT/LINE/TEXT P/L/T:

Reply: P (specify point)

L (specify straight line from last recorded point to the point about to be identified)

T (specify text at the point about to be identified)

*** PLUTARCH SYSTEM ***

DIAGRAMS

HISTPIE

MAPS

INSTPLOT

PERM

INPUTS

LEGENDS

Figure 7.1 Primary Tree (Global Segment Menu)

Teletype message: X COORDINATE:

Reply: decimal number in range 0 - 1023 terminated by full stop.

Teletype message: Y COORDINATE:

Reply: decimal number in range 0 - 1023 terminated by full stop.

The graduated border is removed, then a point or line is generated as required, or the Legend Segment (see paragraph 7.1.6 below) is entered for text generation as appropriate.

- D Delete whole display and initialise, displaying the tracking cross.
- F Set Frame marker. This causes all lines generated subsequently to be truly vertical or horizontal. The position of the light-pen indicates only one coordinate of the endpoint of the new line, the other being determined by the horizontal or vertical property. If a predefined point is indicated as the endpoint of the new line the previous line and the new line are adjusted to give a true right angle.
- H Horizontal A proportions frame.
- M Multidimensional Scaling plot. The input is a paper tape with the following data:

symbol type: decimal digit in range 0-4

- 0 terminate
- 1 +
- 2 diamond
- 3 X
- 4 dot

any number of pairs of X and Y coordinates from MDSCAL,
terminated by the single value 101;

alphanumeric labels for all points, each terminated by *;

option mode: decimal digit in range 0-2

- 0 terminate
- 1 circle mode
- 2 linkage mode;

if the circle mode is specified, there follow the radii of the circles to be circumscribed about the points in the original sequence; in the linkage mode, 0 starts a new chain consisting of the points specified (by original sequence numbers), while - 1 terminates linkage mode. Any number of modes may be called.

The result of this option is a scalogram with labelled points which may have circles (to indicate the importance of each group) or linkage (to indicate relationship between groups as a minimum spanning tree) or both added. The MDSCAL coordinates are scaled without distortion to fit the vertical A proportions frame on the screen.

N Reset frame marker to Normal state. Lines are subsequently drawn at any orientation, the lightpen indicating both X and Y coordinates of the endpoint of a new line.

O Draw circle circumscribed about last specified point.

Teletype message: RADIUS:

Reply: decimal number in range 0 - 512 terminated by full stop.

A circle is generated with centre the last specified point and radius as specified.

P Phenon (dendrogram) plot. The input is a paper tape with the following data:

number of branches in the dendrogram;

numerical identifiers of the branches in sequence left to right;

alphanumeric labels for the branches in sequence left to right, each terminated by *;

any number of sets of:

phenon percentage (a percentage value of - 1 causes termination);

2 numerical identifiers of the two groups
to be joined;

The dendrogram is drawn as required. A percentage scale is drawn at the left hand side, and groups are labelled at their head. The whole diagram is scaled to fit the horizontal A proportions frame. Two or more groups may join at any percentage level.

R Reset frame hold delay for sketch option.

Teletype message: FRAME HOLDS:

Reply: number of frame holds required in sketch delay terminated by full stop.

S Sketch. The routine switches the tracking cross on, then waits for a hardware function key to be pressed. This is the signal that the lightpen is positioned at the start of the desired curve. Thereafter the routine waits until the modulus of the x change or modulus of the y change exceeds a specified value, then draws a small straight-line vector to the new position, adding the vector to the old buffer. Next the routine waits for the specified number of frame holds (this number can be adjusted using the R facility above to make the time delay shorter or longer and hence effectively make the vector lengths finer or coarser), then deletes the curve, returning to the point where the moduli of the x and y changes are checked. The curve continues to be extended until a hardware function key is pressed before the specified number of frame holds occur. This is taken as the signal for the end of the sketched curve.

T Switch Tracking Cross on

U Switch Tracking Cross off (Untrack). This is necessary for the final plotting of the diagram from the display file if the

tracking cross is not to appear in the hard copy.

X EXit to global segment

7.1.1.2 *Function Key options*

- 1 Specify point. The coordinates of the tracking cross are taken as the defined point, which is added to the point list.
- 2 Retrieve coordinates of previously defined point. The numerical identifiers of all the points in the point list are displayed in their correct positions by the procedure PTNUMS. The desired point may be indicated by pointing the activated lightpen at the corresponding identifier, or by typing the identifier on the teletype and terminating by a full stop (this option is useful if the identifier is obscured by other features).
- 3 Draw line from last specified point to current position of tracking cross. If the frame marker is reset the line is drawn direct, otherwise a truly horizontal or vertical line is drawn depending on the relative magnitudes of the x and y changes (in this case the x coordinate is determined by the previous point and the y coordinate by the light pen, or vice versa). The end of the new line is added to the point list in all cases, and this point then becomes the current point.
- 4 Draw line from last specified point to previously defined point. The numerical identifiers of all points in the point list are displayed in their correct positions by the procedure PTNUMS. The desired termination point for the new line may be indicated by pointing the activated lightpen at the corresponding identifier, or by typing the identifier on the teletype and terminating by a full stop (this option is useful if the identifier is obscured by other features). If the frame

marker is reset the line is drawn direct, otherwise a truly horizontal or vertical line is drawn depending on the relative magnitudes of the x and y changes (this involves repositioning the previous point and redrawing the previous line; the coordinates of the previous point are also altered in the point list). The end of the new line becomes the current point, but since it is already in the point list it is not added to this list.

5 Delete item indicated by the lightpen.

6 Position text in desired location.

Teletype message: PREV POINT Y/N:

Reply: Y Retrieve coordinates of previously defined point for positioning the text. The numerical identifiers of all the points in the point list are displayed in their correct positions by the procedure PTNUMS. The desired point may be indicated by pointing the activated lightpen at the corresponding identifier, or by typing the identifier on the teletype and terminating by a full stop (this option is useful if the identifier is obscured by other features).

N Take coordinates of the tracking cross as the position of the text.

In both cases the Legends Segment is entered for specification of the text (see paragraph 7.1.6 below).

7 Enter facilities for 3D Rotation and Perspective. The input is a paper tape with the following data:

Any number of point and line specifications of
the format

P

northings	}	from the origin
eastings		
depth		

alphanumeric label of up to 10 characters terminated by *
(for a point), or

L

northings	}	from the origin
eastings		
depth		

(for a line to be drawn from the last specified point or
endpoint to a specified endpoint).

E terminates the specifications.

When the character E is detected the diagram specified by the
points and lines is drawn in simulated 3D, with perspective.
At each stage the specified point or endpoint of a line becomes
the new current point (i.e. starting point for a subsequent
line). The "hidden line" problem is avoided in the case of
archaeological block diagrams by displaying the front, top and
one side view only and rotating the figure so that the back,
bottom and other side view would in any case be hidden. By
this simple artifice the impression of true hidden line removal
may be given. The screen x coordinates are set to the
northings, y coordinates to the negatives of the depths and z
coordinates to the eastings (out of the screen positive).

The light button cycle is then entered. All z coordi-
nates are monitored, and if any is greater than or equal to

the observer's position (which is initialised at 30 inches from the screen for perspective calculations), the observer's position is moved further out by $\frac{1}{2}$ inch, i.e. the observer is not allowed to move inside the displayed object. Each point is displayed in its simulated perspective position (using procedure PERS) unless the perspective marker is zero when perspective calculations are not performed - this feature is useful in the display of sections, when it is desired that no distortion shall take place; moreover, in the case of cuboidal blocks of a diagrammatic excavation, the remainder of the sides of the block are hidden behind the boundary lines of the section itself and are not apparent in the section displayed. Labels for the points are added in small characters. Lines are drawn from the previous coordinates to the current coordinates, each in perspective (unless the perspective marker is zero) and the illusion of depth is enhanced by brightening lines which are near the observer and dimming those which are far away. The lines are drawn at dim, normal or bright illumination depending on whether the line is wholly behind, intersects or is wholly in front of the screen plane in the simulation. A series of light buttons is displayed along the bottom of the screen as follows:

XYROT YZROT ZXROT TCIN TCOU PNRT PNLFT ZX+90 STOP

One or more of the light buttons may be selected by the activated lightpen with the following actions:

XYROT	Rotate about Z axis 5°
YZROT	Rotate about X axis 5°
ZXROT	Rotate about Y axis 5°
TCIN	Track observer in $\frac{1}{2}$ inch
TCOUT	Track observer out $\frac{1}{2}$ inch
PNRT	Pan view to right (object moves left) $\frac{1}{2}$ inch
PNLFT	Pan view to left (object moves right) $\frac{1}{2}$ inch
ZX+90	Rotate about Y axis 90° (for side section)
STOP	Remove the light buttons so that hard copy may be taken, and await a function key depression.

All light buttons other than ZX+90 and STOP allow the cycle to be repeated as desired. During the normal light button cycle the light buttons are not continually written into the display file, as is the modified figure. ZX+90 and STOP cause the light buttons to be deleted, and a wait loop is entered where the depression of a function key is awaited, as follows:

- 1 Restore light button mode
- 2 Read a new input tape
- 3 Sketch option. Await the depression of a hardware function key. This is the signal that the lightpen is positioned at the start of the desired curve. Thereafter the routine waits until the modulus of the x change or modulus of the y change exceeds a specified value, then draws a small straight-line vector to the new position, adding the vector to the old buffer. Next the routine waits for the specified number of frame holds (this number can be adjusted using function key 7) then deletes the curve, returning to the point where the moduli of the x and y changes are checked. The curve continues to be extended until a hardware function key is pressed before the specified number of frame holds occur. This is taken as the signal for the end of the sketched curve.
- 4 Reverse sense of perspective marker, i.e. remove perspective if present, or display perspective if not present. Enter the light button mode.
- 5 Delete item indicated by lightpen.
- 6 Position text at point indicated by lightpen. This option enters the Legends Segment for text specification (see paragraph 7.1.6 below).

7 Reset frame hold delay for sketch option.

Teletype message: FRAME HOLDS:

Reply: number of frame holds required in sketch delay terminated by full stop.

8 Exit from 3D Rotation and Perspective option to Diagrams Segment options.

Since the lightpen tracking routine is not normally operative (it is switched on specially for sketching and text), it is necessary to enter a waiting loop whenever keys 3 or 6 are pressed. The tracking cross is then moved to the desired starting position using the lightpen, and this is indicated by pressing any function key. In the sketch option the tracking routine remains active, sketching a line until a function key is depressed, when the tracking cross is removed. Function keys 1, 2 and 4 cause entry to the light button mode, while function keys 3, 5 and 7 remain in the key setting mode. Function key 6 enters the Legend Segment and function key 8 causes exit to the Diagrams Segment control.

8 (Diagrams Segment Control)

Read pot profile specified by d-Mac coordinates. The input is a paper tape in the following format:

pot identifier between string quotes;

scale factor e.g. 4.0 means quarter scale;

d-Mac coordinates for the bottom and top of the centre line of the pot;

X0000 Y0000

d-Mac coordinates of the inner profile, bottom to top;

X0001 Y0001

X0000 Y0000

d-Mac coordinates of the outer profile, top to bottom;

X0001 Y0001

halt code

The pot diagram is rotated within the computer until truly vertical, smoothed and scaled to standard size. The pot is then drawn actual size in conventional left-hand section and right-hand elevation with suppression of re-entrant footrim. Vertical centre line and horizontal rim line are added. The pot is drawn on the display using procedure DDRAWF. A title is added with the scale of the drawing (initially 1.0).

Teletype Message: SCALE:

Reply: required scale terminated by full stop (e.g. 25. gives $\frac{1}{4}$ scale, 50. gives $\frac{1}{2}$ scale etc.).

The message is repeated, allowing various scales to be tried until satisfactory. This is indicated by typing 0. in reply to the scale message.

7.1.2 Histogram and Piechart Segment

The loading of this segment is indicated by display of the following options:

```
** HISTPIE OPTIONS **  
  
LAZYHIST  
  
LAZYP  
IE  
  
HISTOGRAM  
  
PIECHART  
  
POLLEN  
  
GLOBAL
```

The required option is selected by pointing the activated lightpen at the corresponding item in the menu. The command is acknowledged by the appearance of a bright-up border around the item selected.

The options available are described below.

7.1.2.1 LAZYHIST (Automatic histogram)

This option expects that data has already been loaded into the global arrays, usually by means of the Input Segment (see paragraph 7.1.7.2). The procedure PREHIST is used to find the range of the items and to calculate the number and size of the histogram steps. The procedure HISTOGRAM is used to plot the histogram on the screen. The Diagrams Segment is then loaded for the addition of borders, legends etc. (see Figure 7.2).

7.1.2.2 LAZYPIC (Automatic piechart)

This option expects that data has already been loaded into the global arrays, usually by means of the Input Segment (see paragraph 7.1.7.2). The procedure PREHIST is used to find the range of the items and to calculate the number and size of the histogram steps. The procedure HISTOGRAM is used without plotting to calculate the counts of items in each interval, then the procedure PIECHART plots the piechart on the screen. The Diagrams Segment is then loaded for the addition of borders, legends etc. (see Figure 7.3).

7.1.2.3 HISTOGRAM

This option plots a histogram for which the range, number of steps and interval have to be specified. The Diagrams Segment is then loaded.

7.1.2.4 PIECHART

This option uses the procedure HISTOGRAM without plotting to calculate the counts of items in each interval. The range, number of steps and interval have to be specified. The procedure PIECHART plots the piechart on the screen. The Diagrams Segment is then loaded.

7.1.2.5 POLLEN

This option expects that pollen data has already been loaded

POTTERS FIELDS 1965

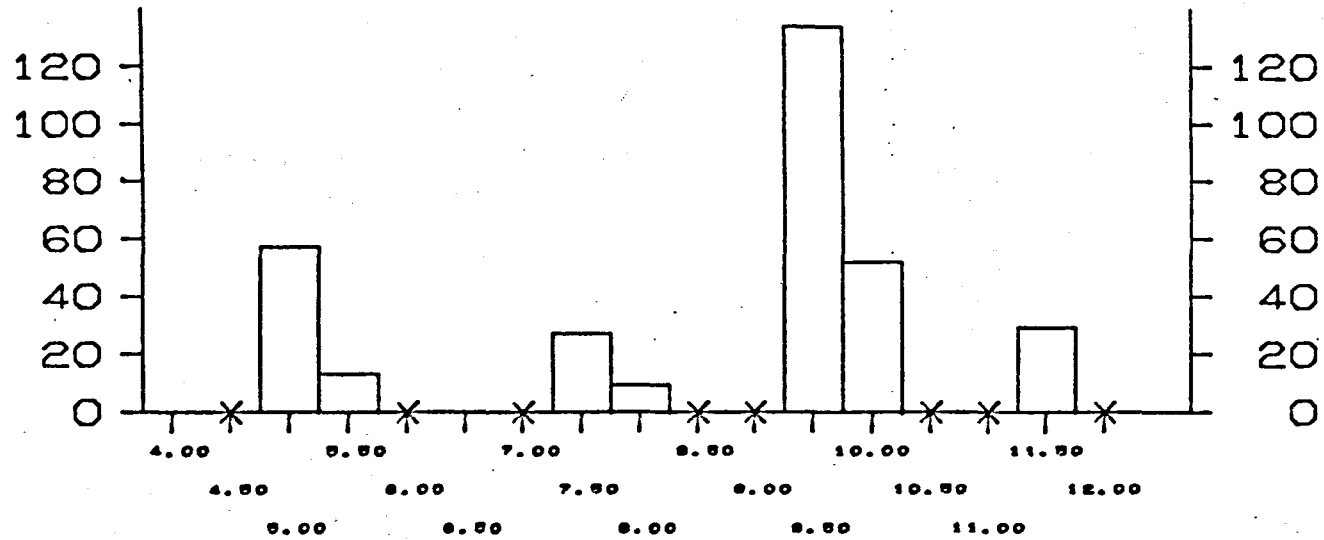
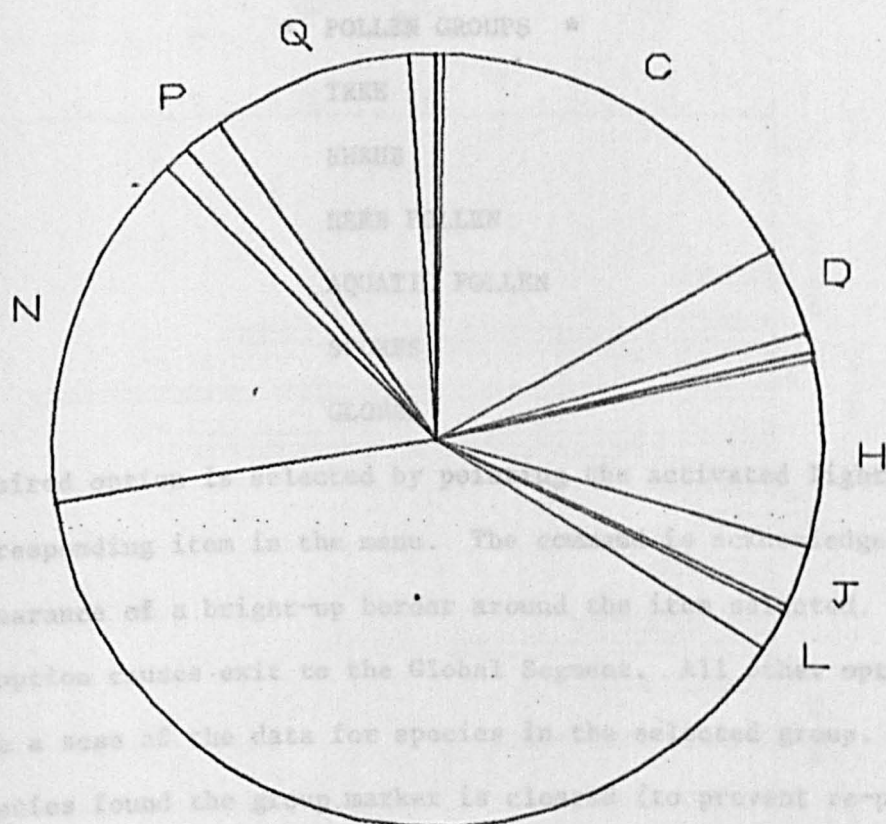


Figure 7.2 Automatic Histogram

POTTERS FIELDS 1965



	M	
C	5.00	16.43 %
D	5.50	3.75 %
H	7.50	7.78 %
J	8.00	2.59 %
L	9.00	1.73 %
M	9.50	38.62 %
N	10.00	14.99 %
P	11.00	1.73 %
Q	11.50	8.36 %

Figure 7.3 Automatic Piechart

into the global arrays, usually by means of the Input Segment (see paragraph 7.1.7.1). The routine performs statistical calculations on the pollen data and outputs the results in histogram form. The following pollen group options are displayed:

* POLLEN GROUPS *

TREE

SHRUB

HERB POLLEN

AQUATIC POLLEN

SPORES

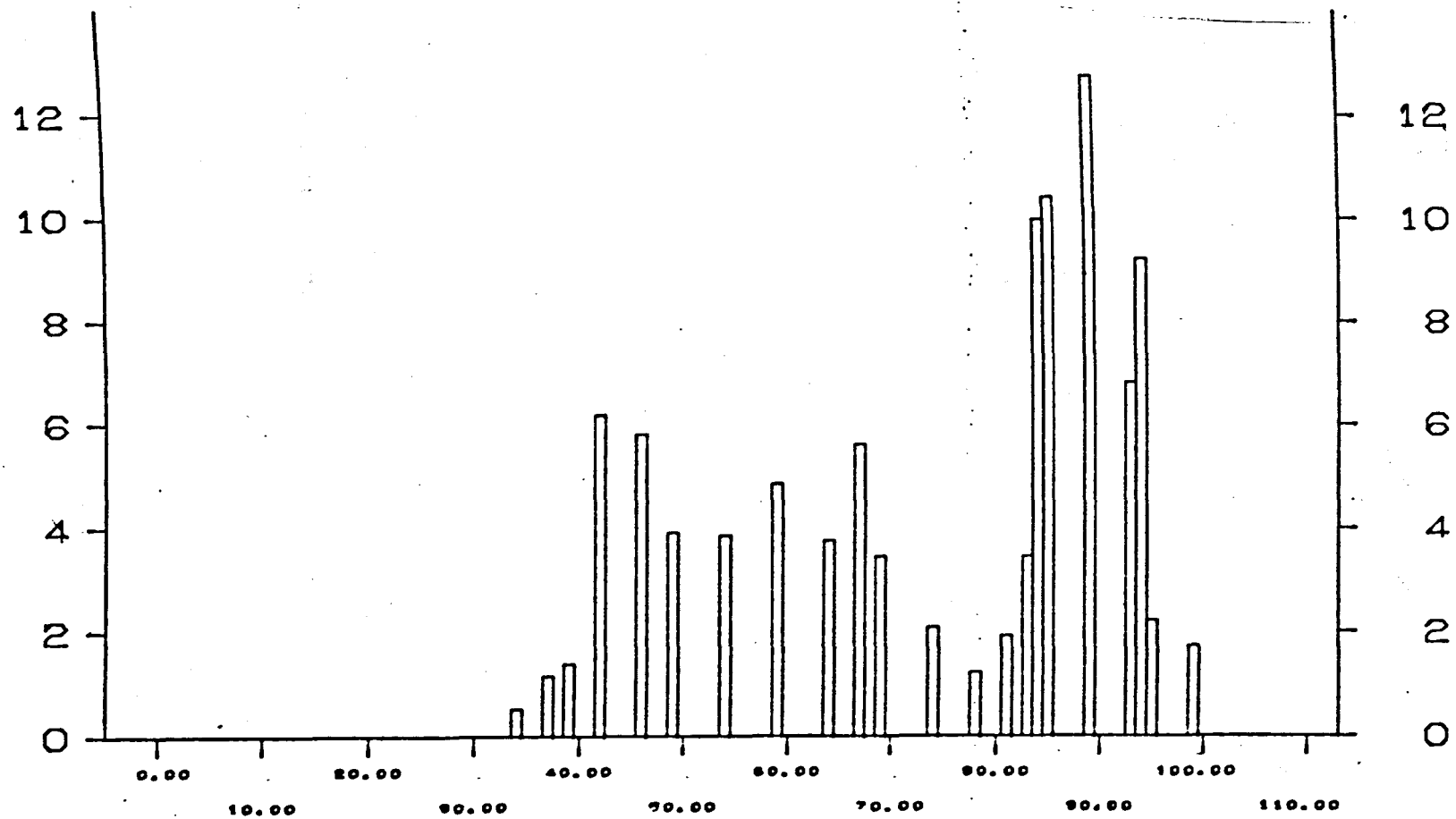
GLOBAL

The required option is selected by pointing the activated lightpen at the corresponding item in the menu. The command is acknowledged by the appearance of a bright-up border around the item selected. The GLOBAL option causes exit to the Global Segment. All other options initiate a scan of the data for species in the selected group. For each species found the group marker is cleared (to prevent re-processing of the data on subsequent calls) and the percentages of the group pollen count and total pollen count are calculated for every sampling depth. The results are printed in a table headed by the species name. The message

*** ALL SPECIES BELOW ARE EXCLUDED FROM POLLEN SUM

is output before the shrubs (for post-glacial studies) or before the aquatic pollen (for late-glacial studies), i.e. only the trees are included in the total pollen count for post-glacial studies, and only the trees, shrubs and herb pollen for late-glacial studies. Species with insignificant pollen counts for all depths are not displayed and the processing of the next species ensues immediately, otherwise the pollen diagram is displayed (see Figure 7.4) and the program awaits

RUMEX



GLANLLYNNAU, CAERNARVONSHIRE

LATE GLACIAL

Figure 7.4

Pollen Diagram

the depression of a function key, or a command to plot the display file. The dimensions of the diagram are calculated automatically, and a depth scale, percentage of total pollen count scale, the site name, post-glacial or late-glacial type, and species name are added. The actions of the function keys on display of a pollen diagram are as follows:

- 1 Continue with group processing
- 2 Enter Diagrams Segment for the addition of borders etc.
- 3-8 As for 1

If at any time the species in the selected group are exhausted, control is returned to the place in the program where the group options are displayed, allowing selection of another group or exit to the Global Segment.

7.1.3 *Maps Segment*

7.1.3.1

The loading of this segment direct from the Global Segment causes immediate loading of the Input Segment (maps input option, see paragraph 7.1.7.6).

7.1.3.2

The automatic loading of this Segment by the Input Segment causes a map to be drawn on the display using data read from magnetic tape. The extreme x and y coordinates of the first part of the map read from magnetic tape are used in the positioning of the map on the display. A scale, grid north point and border are added automatically (see Figure 7.5). If the distribution plot option was selected in the Input Segment then the grid references placed in the global arrays by the Input Segment are plotted using specified symbols (see Figure 7.6). Finally the Diagrams Segment is loaded for the addition of titles etc.

Staffordshire

GRID
NORTH



10 KM

Figure 7.5 Map of Staffordshire without Distribution Plot

Roman Britain

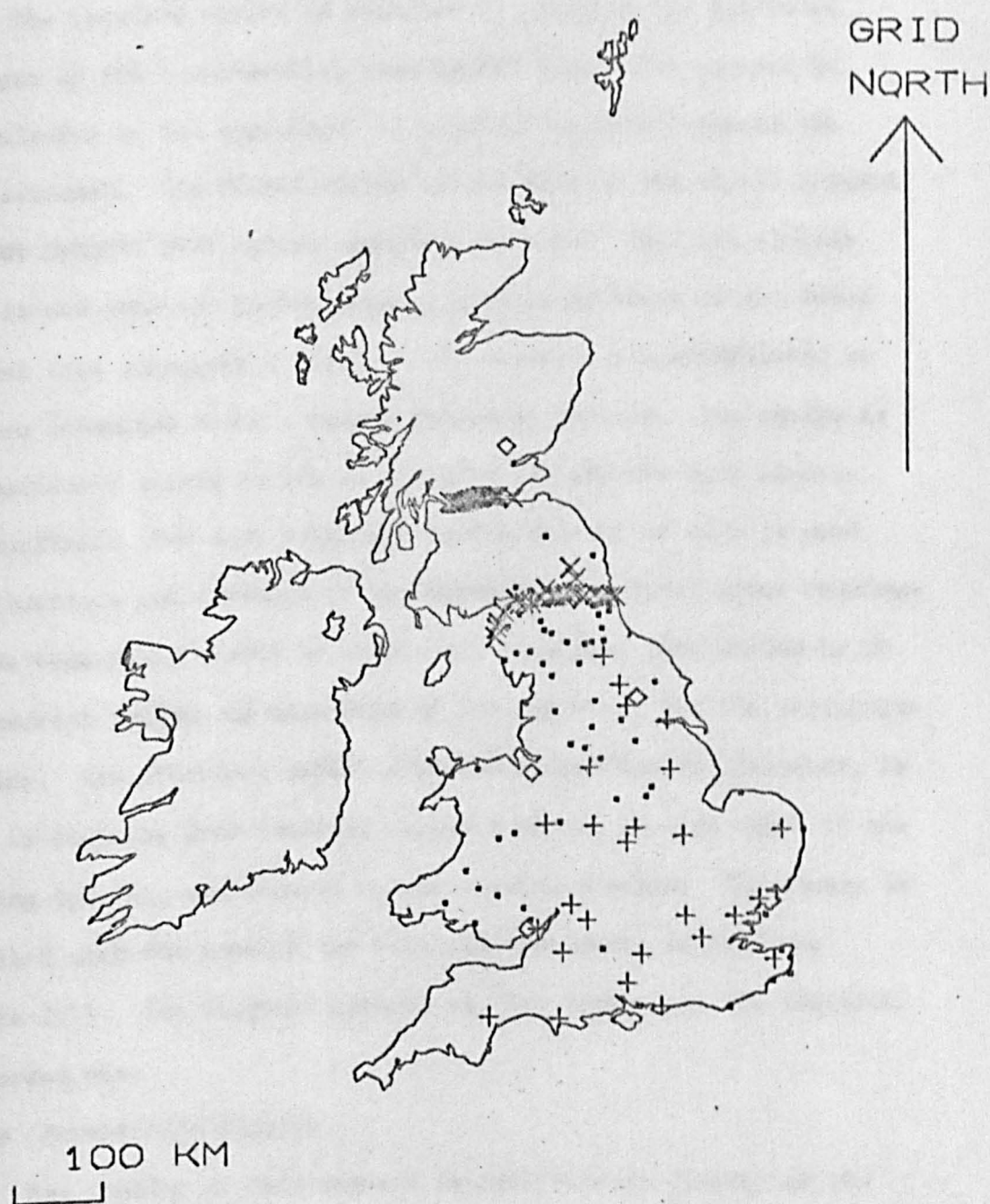


Figure 7.6 Map of Roman Britain with Distribution Plot

7.1.4 Instrument Plot Segment

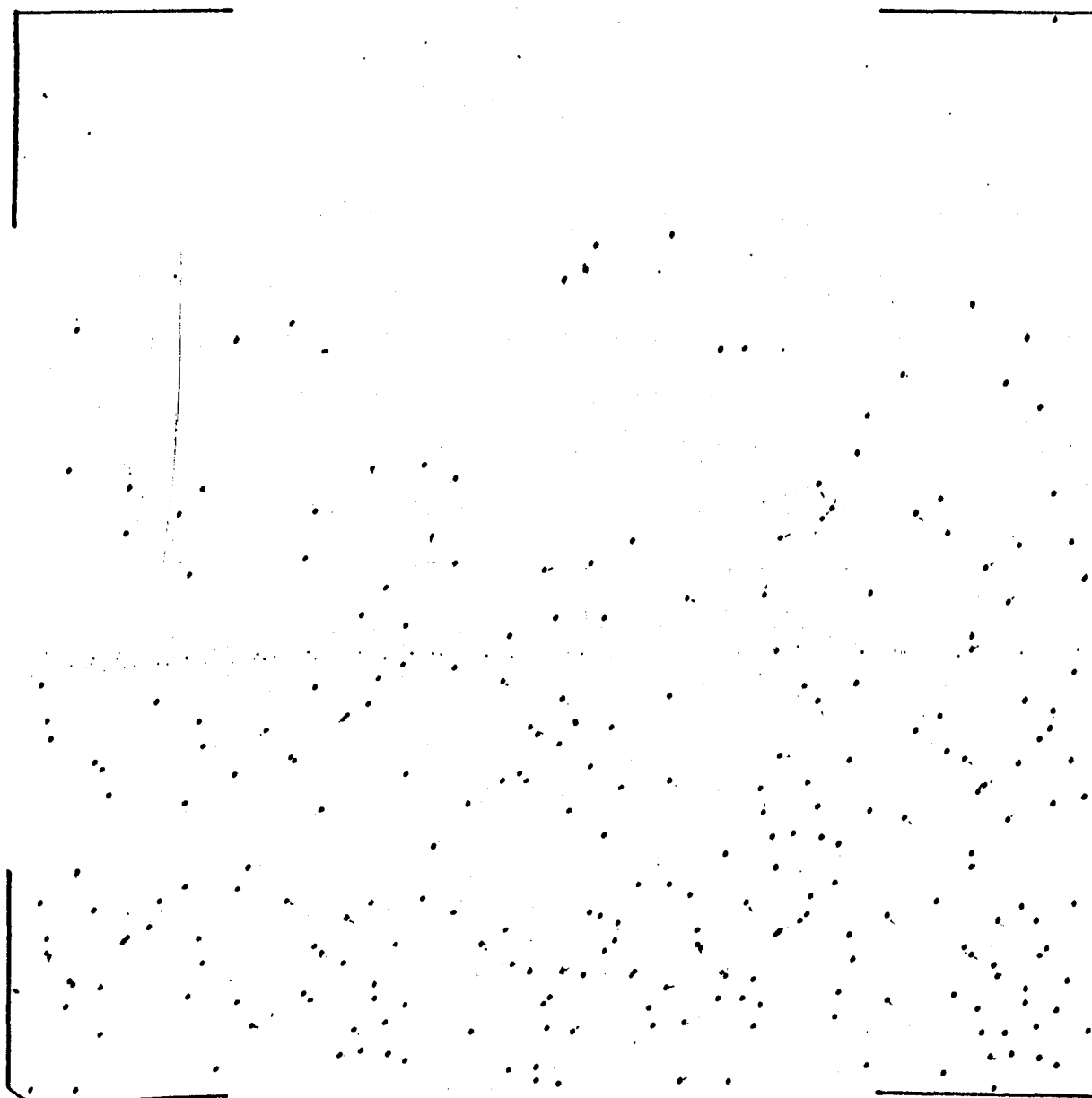
The loading of this segment is indicated by display of the following options:

```
    ** INSTRUMENT PLOT OPTIONS **  
  
    DOT DENSITY PLOT  
  
    GLOBAL
```

The required option is selected by pointing the activated lightpen at the corresponding item in the menu. The command is acknowledged by the appearance of a bright-up border around the item selected. The GLOBAL option causes exit to the Global Segment. The DOT DENSITY PLOT option expects that survey data has already been loaded into the global arrays, usually by means of the Input Segment (see paragraph 7.1.7.5). The readings are manipulated to enhance anomalies using a simple *filtering* process. The square is automatically scaled to fit on the display, and the four corners are outlined. For each reading a random scatter of dots is used to illustrate the strength of the anomaly. For Banjo meter readings it has been found useful to square all readings, then divide by 20. The nearest integer is then used as the dot count for the particular reading. The procedure RANNO, a pseudo-random number generator, is used to position dots randomly within a square of side equal to the reading spacing, and centred on the reading station. The square is labelled with the name of the site and the square number (see Figure 7.7). The Diagrams Segment is then loaded for the addition of border, etc.

7.1.5 Permutation Segment

The loading of this segment is indicated by display of the following options:



SOUTH CADBURY

SQUARE 23

Figure 7.7 Instrument Survey Plot

**** PERMUTATION OPTIONS ****

POTTERY

CLOSED FIGURES

GLOBAL

The required option is selected by pointing the activated lightpen at the corresponding item in the menu. The command is acknowledged by the appearance of a bright-up border around the item selected. The GLOBAL option causes exit to the Global Segment. Other options expect that profile data for the required number of artefacts (pots or closed figures, such as projectile points) has already been loaded into the global arrays, usually by means of the Input Segment (see paragraphs 7.1.7.3 and 7.1.7.4). The recursive procedure PERM is used to permute the figures on the display, with the intention of allowing the operator to choose the most suitable evolutionary sequence. After each permutation is displayed (see Figures 7.8 and 7.9), the program awaits the depression of a hardware function key as follows:

- 1 Display next permutation
- 2 Load Diagrams Segment for the addition of border,
 legends, etc.
- 3-8 As for function key 1.

If the permutation is exhausted the Global Segment is entered.

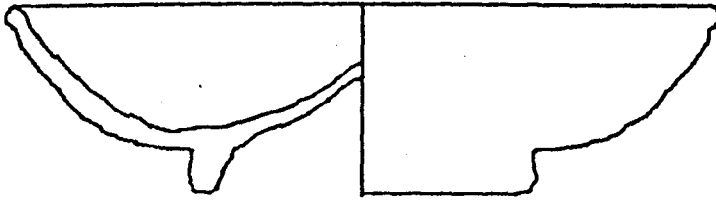
7.1.6 *Legends Segment*

This segment may be entered direct from the Global Segment or via function key 6 from the Diagrams Segment (see paragraph 7.1.1.2). The loading of the segment is indicated by the teletype message:

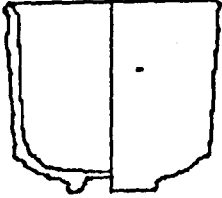
FONT:

Reply: letter A (for the system font)

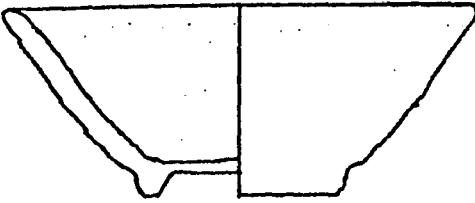
 or letter B (for the PLUTARCH font which resembles OCRB)



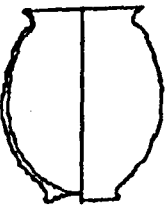
1



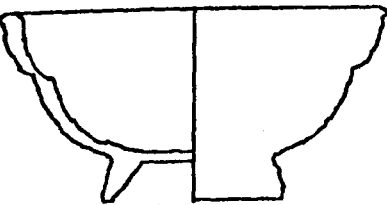
2



5



3



4

Figure 7.8 Permutation 12534 of five pottery profiles



1



3



2



5



4

Figure 7.9 Permutation 1325⁴ of five closed figures
(projectile points)

followed by a single size digit:

- 1 approx. 12 characters/inch
- 2 approx. 6 characters/inch
- 3 approx. 3 characters/inch

followed (in the case of the OCRB font) by an italic digit:

- 0 No italics
- 1 Italic slope of approx. 11°
- 2 Italic slope of approx. 22°
- 3 Italic slope of approx. 31°

and an orientation digit:

- 0 Horizontal (left to right)
- 1 Vertical (bottom to top)
- 2 Horizontal (upside down, right to left)
- 3 Vertical (top to bottom)

Teletype message: TEXT:

Reply: required text terminated by full stop.

The following characters are used as control symbols and hence may not be displayed as part of the text:

l Lower case - all characters are output as lower case until a further control symbol is input

\$ Upper case - all characters are output as upper case until a further control symbol is input

(full stop) Termination of text.

In the absence of a shift control symbol the text is assumed to be upper case. After each line of text has been displayed the program awaits the depression of a hardware function key as follows:

1-5 Load Diagrams Segment for the addition of border, etc.

6 Accept another line of text (output of FONT message)

7,8 As for function keys 1-5

Subsequent lines of text are automatically spaced according to the height of the required font. The first line of text is positioned starting at the top left-hand corner of the screen (for direct entry

from the Global Segment) or by the lightpen coordinates (for entry via function key 6 in the Diagrams Segment, see paragraph 7.1.1.2). Samples of typefaces are shown in Figures 7.10 and 7.11.

7.1.7 *Inputs Segment*

The loading of this segment is indicated by display of the following options:

**** INPUT OPTIONS ****

POLLEN

LAZY

READFIG

READCLO

INSTPLOT

MAPS

GLOBAL

This segment must be entered before several of the other segments are entered in order to load the global arrays with data. The required option is selected by pointing the activated lightpen at the corresponding item in the menu. The command is acknowledged by the appearance of a bright-up border around the item selected and the magnetic tape library is displayed. The GLOBAL option causes exit to the Global Segment. Other options are described below.

7.1.7.1 *Pollen data*

This option must be used before the Pollen option of the Histpie Segment is loaded (see paragraph 7.1.2.5). The required inputs are a paper tape and the pollen species master magnetic tape. The paper tape has the following data:

Number of pollen samples taken (≤ 25);

Name of site between string quotes;

This is a Message
ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
0123456789

"%& () * = - ; + / . < > ,
UP down up
UP down UP
Message bracketed

Figure 7.10 Samples of the system font (size 2)
as made available by the provision
of shift control symbols

SMALLER MESSAGE
EVEN SMALLER MESSAGE

This is a & Message

Demonstration of OCRB

nnnnnnnnnnnn

mmmmmmmmmmmm

IIIIIIII

IIIIIIIIII

ffffffffffff

jjjjjjjjjj

kkkkkkkkkk

llllllllll

tttttttttt

ABCDEFGHIJKLMNOPQRSTUVWXYZ01223456789

abcdefghijklmnopqrstuvwxyz

Figure 7.11 Samples of the system and OCRB fonts.

- Note a) the common alignment of the different sizes of system font
b) the variable shift and italic slope of the OCRB font
c) the differing pitches of the characters:
i (2 units)
I,j,l (4 units)
f,k,t (5 units)
most other characters (6 units)
m (8 units)

Period character:

L Late-Glacial

P Post-Glacial;

Depths of all the pollen samples, in order from the surface down;

Any number of sets of data as follows:

Species name terminated by * (max. 29 characters);
pollen counts for each sampling depth in order from
the surface down;

* (in place of the first character of a species name) to
terminate the paper tape input.

The pollen species master magnetic tape is called by a bright-up border around the correct tape in the displayed tape library and by the teletype message:

MOUNT INDICATED TAPE ON ALLOCATED HANDLER

The required information is displayed in the format

< tape identifier > reel no. handler (contents)

and is outlined by a bright-up border. When the correct tape has been mounted on the allocated handler, reply . (full stop) on the teletype.

All species names input from paper tape are checked from magnetic tape and allocated to their pollen group (trees, shrubs, herb pollen, aquatic pollen or spores). As each species is read from magnetic tape the species name is printed and the input list is scanned. During this scan, species which have already been allocated to groups are ignored, and if the species is located in the input list the message PRESENT is appended to the species name on the printer, and a group code (based on the second letter of

the pollen group, which gives a unique code) is allocated to the species in the input list. The pollen counts for each depth are added to the corresponding pollen group sums, then the program returns to read a new pollen species from magnetic tape. When the magnetic tape scan is complete, and if all species in the input list have been recognised, the correct total pollen sums for all depths (all trees for post-glacial studies, or all trees, shrubs and herb pollen for late-glacial studies) are calculated. The histogram increment for the depth scale is calculated automatically, then the Histogram and Piechart Segment is loaded automatically and entered at the Pollen option (see paragraph 7.1.2.5 above).

7.1.7.2 *Lazy Input*

This option must be used before the LAZYHIST and LAZYPIC options of the Histpie Segment are loaded (see paragraphs 7.1.2.1 and 7.1.2.2). It is a mode of input which reads up to 1000 real numbers from paper tape and loads them into the global arrays. The numbers must be less than 1000000, as a number greater than or equal to 1000000 is used as a terminator. The numbers need not be counted before input: the routine automatically counts the items, lists them on the printer, then sets the first item in the global array equal to the count before entering the Global Segment.

7.1.7.3 *Pot Profile Input (READFIG)*

This option must be used before the Permutation Segment is loaded if pottery diagrams are to be permuted (see paragraph 7.1.5).

Teletype message: NUMBER OF FIGURES 1-5:

Reply: Single digit in the range 1-5

The required number of pottery profiles are read and stored in the global arrays. The profiles are specified by d-Mac coordinates.

The input is a paper tape in the following format:

pot identifier between string quotes;

scale factor e.g. 4.0 means quarter scale;

d-Mac coordinates for the bottom and top of the centre
line of the pot;

X0000 Y0000

d-Mac coordinates of the inner profile, bottom to top;

X0001 Y0001

X0000 Y0000

d-Mac coordinates of the outer profile, top to bottom;

X0001 Y0001

halt code .

The pot diagram is rotated within the computer until truly vertical, smoothed and scaled to standard size, with adjustment of the scale factor, then stored in the global arrays. When all required pot profiles have been input the Global Segment is entered.

7.1.7.4 Closed Figure Input (READCLO)

This option is used to input closed figures, e.g. projectile point outlines, and must be used before the Permutation Segment is loaded if closed figures are to be permuted (see paragraph 7.1.5).

Teletype message: NUMBER OF CLOSED FIGURES 1-5:

Reply: Single digit in the range 1-5

The required number of closed figures are read and stored in the global arrays. The figures are specified by d-Mac coordinates.

The input is a paper tape in the following format:

identifier between string quotes;

scale factor e.g. 4.0 means quarter scale;

X0000 Y0000

d-Mac coordinates of the closed figure, starting and finishing at an arbitrary point preferably at the same level and to the right of the approximate position of the centroid of the closed figure;

X0001 Y0001

halt code .

The figure is smoothed, then standardised by finding its areal centroid and the point on the profile furthest away from this centroid (usually the pointed end for a projectile point). The whole figure is then rotated so that this furthest point is vertically above the centroid, and the distance from this point is standardised, with adjustment of the scale factor. The area is found as the summation of a number of elementary rectangular strips as the curve is followed from beginning to end, some of the strips being added to the summed area and some subtracted until the circuit is complete. The centroid is calculated during the same pass. During the rotation and standardisation pass each d-Mac coordinate has a polar angle allocated to it relative to the centroid, and there is a check that the polar angles always increase around the profile. The standardised coordinates are stored in the global arrays. When all requested closed figures have been input the Global Segment is entered.

7.1.7.5 *Instrument Plot Input*

This option must be used before the Instrument Plot Segment is loaded (see paragraph 7.1.4). The input is a paper tape in the following format:

site identifier between string quotes;

number of readings/line N (< 52);

N sets of:

N readings

line check number of form SSSLL where

SSS is the square number (3 digits)

LL is the line number in range 1 - N (2 digits),

and where the last line has check number 1SSSLL;

The purpose of the line check number is to detect the input of incomplete or oversize lines. Also the LL number is used to determine where the readings shall be stored in the global arrays. The lines are counted during input and when the square terminator 1SSSLL is encountered the total number of lines is checked to be equal to the number of readings/line, i.e. a complete square must be input. The square number is stored in a global variable and the Global Segment is entered.

7.1.7.6 *Maps Input*

This option is automatically loaded by the Maps Segment (see paragraph 7.1.3.1) or may be entered direct from the Global Segment using menu commands INPUTS (see paragraph 7.1) and MAPS (see paragraph 7.1.7). In both cases the magnetic tape library is displayed, with the following options:

* MAP OPTIONS *

MAP REFS

OUTLINE ONLY

The required option is selected by pointing the activated lightpen at the corresponding item in the menu. The command is acknowledged by the appearance of a bright-up border around the item selected. The map references option allows a distribution map to be plotted using various symbols, while both options draw the map outline, scale, grid north point and border. A second set of options is

presented as follows:

* MAP OPTIONS *

GBI

STAFFS

These options refer to the required map outline (at present only Great Britain and Ireland (GBI) and Staffordshire (STAFFS) are provided). The required map is selected by pointing the activated lightpen at the corresponding item in the menu. The command is acknowledged by the appearance of a bright-up border around the item selected. The map master magnetic tape is called by a bright-up border around the correct tape in the displayed tape library and by the teletype message:

MOUNT INDICATED TAPE ON ALLOCATED HANDLER

The required information is displayed in the format

< tape identifier > reel no. handler (contents)

and is outlined by a bright-up border. When the correct tape has been mounted on the allocated handler, reply . (full stop) on the teletype.

The magnetic tape is searched for the required map; when this is found the scale is read and transferred to a global variable and the tape is positioned ready for the appropriate map outline coordinates to be read by the Maps Segment (see paragraph 7.1.3.2).

If no grid references are to be plotted, the Maps Segment is loaded (see paragraph 7.1.3.2), otherwise the grid is defined and grid references are read from paper tape in the following format:

(a) LL or L* or **

where LL gives the two-letter code of a single 100 Km

grid square, e.g. SJ for many Staffordshire grid references; or where L* indicates the definition of a whole 500 Km grid square (comprising 25 100 Km grid squares e.g. S* would cover the whole of Southern Britain) - each 100 Km square is allocated a unique code consisting of the first letter of the 500 Km grid square and another letter; ** indicates the definition of a further grid system.

- (b) Following an LL code are two integers specifying the distances in Km east and north respectively of the SW corner of the 100 Km square from the current grid origin; following an L* code are two integers giving similar values for the SW corner of the 500 Km square - the constituent 100 Km squares are allocated coordinates automatically; following an ** code is either the number zero (to terminate the definition phase) or a non-zero real number giving the angle in radians which the base-line of the new grid system makes with the base-line of the first-defined grid system, and two integers giving the east and north displacements in Km of the origin of the new grid system from the origin of the first defined grid system.
- (c) After the definition of grid systems and squares has been terminated by value zero following an ** code, four integers specify the east range (maximum then minimum) and the north range (maximum then minimum) in display units of valid grid references from the origin of the first-defined grid system.
- (d) Any number of plotting symbols and grid references may be specified (symbols must of course be specified before grid references are given, but a defined symbol need only be

specified once and is retained for several grid references, i.e. until another type of symbol is defined). Symbols are defined by code *n where n is a digit in range 1-6 as follows:

- 1 +
- 2 diamond
- 3 X
- 4 dot
- 5 enhanced dot
- 6 doubly-enhanced dot

Grid references are defined by code letters and digits in the format LLEEENNN where LL is a previously-defined 100 Km grid square code, EEE is a 3-digit eastings and NNN a 3-digit northings displacement from the SW corner of the square, given in 100 m units (i.e. the 3 digits represent 10 Km, Km and 100 m values). The code *0 causes termination and the Maps Segment is loaded automatically (see paragraph 7.1.3.2).

7.2 *Description of Method*

Because the code would otherwise not fit into the available store space, the program is segmented so that only frequently-used procedures, variables and arrays are permanently resident together with whatever coding is necessary to perform the job immediately in hand. The store size required is thus the sum of the global procedures etc. and the largest of the other segments. As a consequence of splitting up the coding in this way it is necessary to have available global arrays and variables for the specific purpose of communicating between segments which are not resident in the store at the same time.

The various segments and options are selected by means of the activated lightpen. The option lists are displayed on the screen by means of procedure LTBUT. Each call of this pro-

cedure creates a light button at a defined position on the screen. The position and character length of each light button is stored so that if selected it can be outlined by a bright-up border by procedure OUTLINE in acknowledgment of the command. The border is maintained until 5 frame holds have been received, then the coding for the option is loaded. Successive sets of light buttons are positioned differently so that the lightpen, if it remains activated after a selection, cannot accidentally select an incorrect option from the second set of light buttons at the same point on the screen.

7.2.1 *Other Global Segment Procedures*

The following frequently-used procedures have been placed in the Global Segment; they are fully described in Appendix A:

CHAROUT	Character output complementary to ADVANCE and DECODE.
CHARCLOSE	Closing procedure for use after calls of CHAROUT.
INSBL	Read and check tape mark and SBLO.
INCDI	Read card image or tape mark.
UNPACK	Unpack string to integer form, recognising $\frac{1}{2}$ as case shift to produce lower case and \$ as case normal to produce upper case. The final integer stored is 14 (full stop) in preparation for a call of procedure SCRIPT.
READFIG	Read, rotate, smooth and scale pottery profile.
READCLO	Read, smooth and find centroid and area of a closed figure, e.g. a projectile point. The figure is rotated so that the furthest point from the centroid is vertically above the centroid, then scaled.
DDRAWF	Draw a figure on the display. Rounding errors during scaling procedures are accumulated and corrected when

they amount to one display unit in the x or y direction.

DECIN Read decimal number from teletype.

SCRIPT Produce packed string including lower case alphabets.

½ is recognised as case shift and \$ as case normal. Integer value 14 (full stop) is recognised as termination and causes the closing string quote to be produced.

INCDRAW General display procedure which draws an item (by entering it in the display file) and increments the item count.

SEARCH Check for correspondence between integer strings.

7.2.2 *Primary Tree*

The primary tree is contained in the Global Segment (see paragraph 7.1 and Figure 7.1). It is used to select other segments. A general reset routine is frequently used before display of this tree to delete any existing items remaining on the screen.

7.2.3 *Diagrams Segment*

This segment includes procedure PTNUMS for the display of point numbers in order that previously-defined points may be recalled.

A large number of options are available in this segment, selected by means of hardware function switches and teletype characters. The facilities are fully described in paragraph 7.1.1. Most other segments automatically load this segment for the addition of other features.

7.2.4 *Histogram and Piechart Segment*

This segment includes the procedures PREHIST, HISTOGRAM and PIECHART. All are fully described in Appendix A.

PREHIST Find range of values and calculate histogram
 increment

HISTOGRAM Produce histogram

PIECHART Produce piechart

The facilities of this segment are fully described in paragraph 7.1.2 and illustrated in Figures 7.2 - 7.4.

7.2.5 *Maps Segment*

This segment includes the following procedures, fully described in Appendix A:

DRAWMAP Read map from magnetic tape, draw it on the display and add scale, grid north point and border.

PLOTREFS Read grid references from global arrays and plot them on a map using required symbols.

The segment uses the technique of loading map input automatically. The facilities are described in paragraph 7.1.3 and illustrated in Figures 7.5 and 7.6.

7.2.6 *Instrument Plot Segment*

This segment includes the following procedures, fully described in Appendix A:

FILTER Filtering procedure

RANNO Random number generator

The facilities are described in paragraph 7.1.4 and illustrated in Figure 7.7.

7.2.7 *Permutation Segment*

This segment includes the following procedures:

COORDS Procedure coordinates for pottery profile or closed figure

PERM Recursive permutation of figures on the display

The intention of this segment is to display figures in all possible permutations, allowing the archaeologist to choose the most suitable evolutionary sequence of artefacts. The facilities are described in paragraph 7.1.5 and illustrated in Figures 7.8 and 7.9.

7.2.8 *Legends Segment*

This segment includes the following procedures:

O Produce binary from octal and pack

PREOCRB Define OCRB graphics. This procedure is called once before a series of font specifications

OCRB Plot OCRB font on the display in a manner similar
to procedure OUTSTRING.

The segment allows lines of text to be assembled in differing styles, sizes and orientations. The facilities are fully described in paragraph 7.1.6, and illustrated in Figures 7.10 and 7.11.

7.2.9 *Inputs Segment*

This segment uses procedure CALLMT to outline a specified tape in the displayed magnetic tape library and to await the mounting of the required tape.

The maps input option uses procedure READREFS to define grids and to read grid references and symbol types.

The facilities are fully described in paragraph 7.1.7.

7.3 *Segment Flowcharts*

The relationship and linkage between the segments is shown in Figure 7.12. The size of the blocks is proportional to the actual storage required by the segments.

7.4 *Coding*

The ALGOL 60 coding of the PLUTARCH system with input/output implementation for the ICL 4100 Series is listed in the attached printout.

7.5 *Data Checking Procedures and Error Recovery*

Data vetting is carried out for teletype and paper tape input. In the case of teletype input rejection usually causes the original input demand message to be repeated. For paper tape input either failure messages are output to the line printer or teletype, and/or the whole block is rejected. Failure messages and other error recovery procedures are listed below.

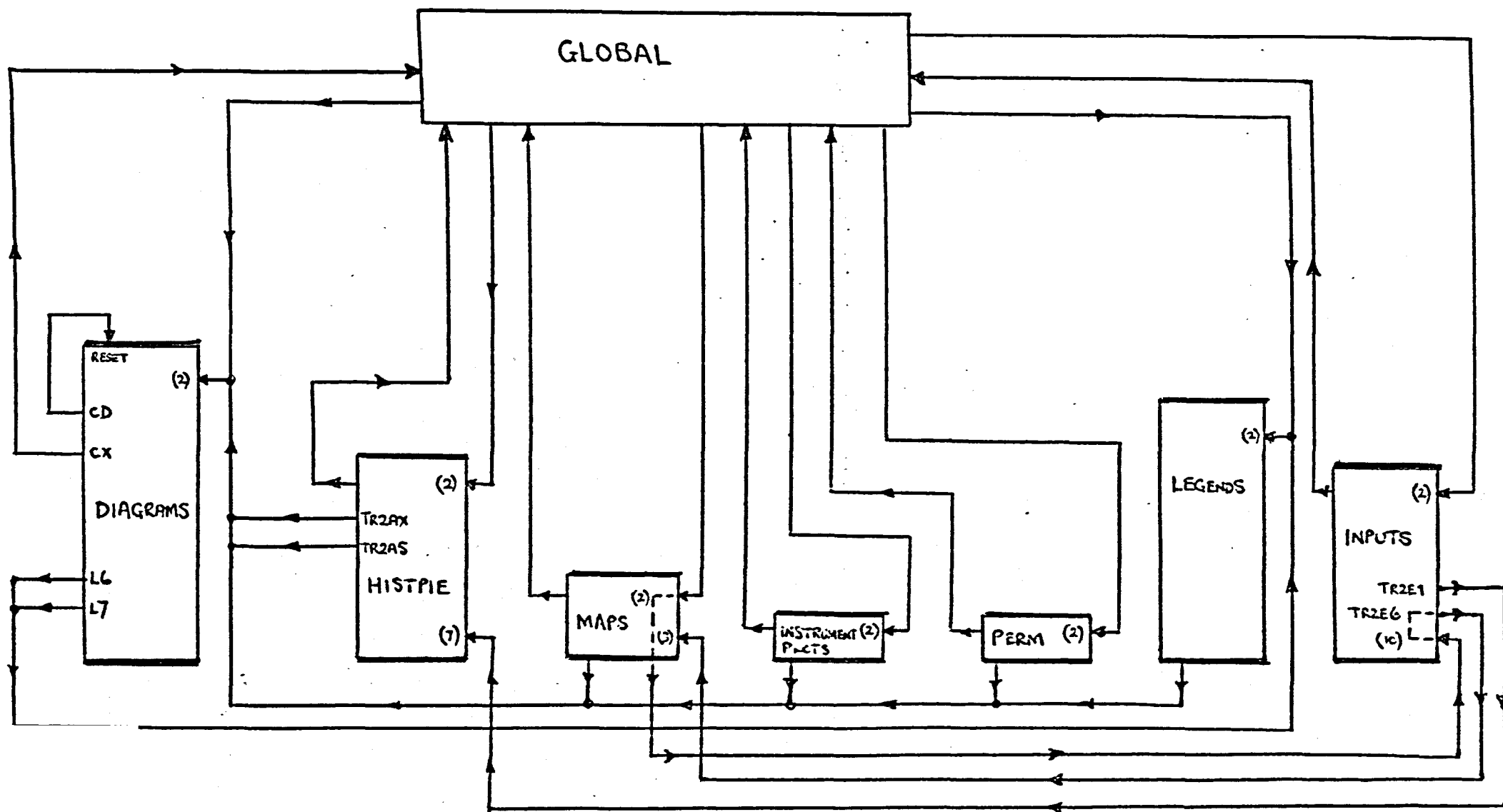


Figure 7.12 Segment relationship and linkage in the PLUTARCH System
The sizes of the blocks represent the relative store sizes taken up by the segments

7.5.1 *Global Segment Error Recovery*

Code Area

Procedure DECIN

Individual digits checked to be in range 0-9 (otherwise rejected). Also checks that at least one valid digit has been input before the terminating . (full stop).

7.5.2 *Diagram Segment Error Recovery*

Code Area

Procedure PTNUMS

Checks that lightpen identifies a point label and not some other item on the screen (continues to await a valid lightpen interrupt).

OR if teletype input is used, individual digits are checked to be in the range 0-9 (otherwise the whole of the number assembled so far is rejected); when the terminating . (full stop) is received the value is checked to be a valid point number (if not the routine enters the general Diagram Segment option select cycle, allowing deletion of the display).

General Option Select Cycle

Teletype character options checked to be a letter or a decimal digit (otherwise ignored).

Option C

- a) Teletype message:
POINT/LINE/TEXT P/L/T:
Reply checked to be P, L or T
(otherwise message is repeated)

b) Teletype message:

X COORDINATE:

Reply checked to be in range 0-1023

(otherwise message is repeated)

c) Teletype message:

Y COORDINATE

Reply checked to be in range 0-1023

(otherwise message is repeated)

Option M

a) Symbol type checked to be non-zero (otherwise general option select cycle is entered).

b) Mode checked to be 1 (circles) or 2 (linkage), otherwise the general option select cycle is entered - this is also used for normal termination.

Option O

Teletype message:

RADIUS:

Reply checked to be in range 0-512 (otherwise message is repeated).

Option P

Number of branches checked to be positive non-zero (otherwise general option select cycle is entered).

Option R

Teletype message:

FRAME HOLDS:

Reply checked to be positive non-zero.

Option 7

Function key 7 in 3D Rotation and Perspective option gives

Teletype message:

FRAME HOLDS:

Reply checked to be positive non-zero.

Option 8

a) If the d-Mac read is invalid the block is ignored and another is read.

b) Teletype message:

SCALE:

Reply checked to be positive non-zero (otherwise the general option select cycle is entered) - this is also used for normal termination.

7.5.3 *Legends Segment Error Recovery*

Code Area

Font Specification

Teletype message:

FONT:

Second character of reply checked to be in range 1-3 (size).

If the first character is B (OCRB) two further characters are read and checked to be in the range 0-3 (italic slope and orientation).

In all cases an error causes the teletype message to be repeated.

7.5.4 *Inputs Segment Error Recovery*

Code Area

Procedure CALLMT

Teletype message:

MOUNT INDICATED TAPE ON ALLOCATED HANDLER

Checks reply is . (full stop), otherwise repeats message.

Pollen input

a) Checks number of samples in range 1-25 (otherwise enters Global Segment)

b) Checks period character is P or L (if a digit is found the Global Segment is entered, otherwise if the character does not equal P or L another character is read).

- c) Checks species names start with a letter
(if not continues to read characters until a letter is found, which is taken as the start of the species name) and are of length 1-29 characters (if not, oversize names are truncated).
- d) Species names which are recognised initiate the printer message PRESENT following the species name.
- e) If a pollen group cannot be allocated to the species the failure message
POLLEN GROUP UNDEFINED
is printed (this is usually an indication that the magnetic tape is corrupt), followed by the name of the species. The Inputs Segment general option select cycle is entered immediately, and further pollen analysis is not possible.
- f) If a species name or names cannot be recognised the message
SPECIES NOT RECOGNISED
is printed followed by a list of the unrecognised names. The Input Segment general option select cycle is entered after all valid species have been processed, but totals and histogram increments have not been calculated and continuation of pollen analysis is not possible.

Code Area

Pot Profiles input

a) Teletype message:

NUMBER OF FIGURES 1-5:

Reply checked to be in range 0-5

(otherwise the message is repeated).

If reply is 0 the Inputs Segment general option select cycle is entered.

b) If the d-Mac read is invalid the block is ignored and another is read.

Closed Figures input

a) Teletype message:

NUMBER OF CLOSED FIGURES 1-5:

Reply checked to be in range 0-5

(otherwise the message is repeated)

If reply is 0 the Inputs Segment general option select cycle is entered.

b) If the d-Mac read is invalid the block is ignored and another is read.

Instrument plot input

a) Number of readings/line checked to be ≤ 51

If not the message

SQUARE TOO BIG

is output to the teletype and the Global Segment is entered.

b) Line terminators checked to have valid line numbers. If not the message

INCORRECT LINE

is output to the teletype and the Global Segment is entered.

c) Total number of lines checked to equal the number of readings/line. If not the message

SQUARE INCOMPLETE

is output to the teletype and the Global

Code Area

Segment is entered.

Maps input

- a) Defined grid squares are checked to start with a valid code (if not characters are read until a valid code is read).
- b) Grid references are checked to start with two valid codes. If not the message
REJECTED INVALID SQUARE CODE
is printed, and the complete grid reference is ignored.
- c) Grid references are checked to have a defined square code. If not the message
REJECTED UNLISTED SQUARE CODE
is printed and the complete grid reference is ignored.
- d) Grid references are checked to have a numerical part ≤ 999999 . If not the message
REJECTED INVALID 6 FIG GR
is printed and the grid reference is ignored.
- e) Grid references are checked to be within the specified eastings and northings range for the required map. If not the message
REJECTED GR OUT OF RANGE
is printed and the grid reference is ignored.
- f) In the search of the magnetic tape if the required map outline is not found the name of the required map is printed followed by the message
NOT FOUND
and the Global Segment is entered.

Appendix A

ALGOL Procedures used in the archaeological programs

This appendix contains a full list of ALGOL procedures which have been developed for this review of the applications of the computer to archaeology. Many of the procedures have been designed for Specific purposes in archaeological analysis. Each is described with particular regard to novel features, and the coding of the most interesting is also included.

For completeness useful procedures in the ICL systems software and EDGAR display language, which are extensions of the ALGOL language in the undefined input/output region, are also included. For these only a brief description is given, but the aim is to show how the original procedures complement the systems procedures. Practical operating details are also given for those systems procedures which are inadequately described in the systems manuals (see especially OUTSTRING and SETP).

This combination of systems procedures and specially-designed procedures is a basis for a flexible ALGOL-based language for the use of archaeologists. A typical archaeological applications package consists of a selection of these procedures built into a skeleton ALGOL program. The full facilities of the ALGOL language as defined in the ALGOL 60 Report (Naur, 1963) are also available. Much use has been made of the specification of instructions applicable to one run of a program only as pseudo-data cards. An example of this is the specification of a Boolean check which is only required for one information retrieval run. The ALGOL statements defining the Boolean check are punched on cards of the colour normally reserved for data, and inserted between labelled points in the program. In this way the facilities of the ALGOL compiler are available to interpret the

Boolean check, which may be as complicated as required. A disadvantage of this mode of working is that the program has to be compiled prior to each run i.e. it is restricted to a batch processing mode, but in many card-based systems this is normal practice in any case for user programs. The main advantage is that the program does not need a complex routine to interpret a variable Boolean check read in true data form at run time.

* * * * *

ACTION (EDGAR)

The most useful of the instructions which hold up the program to await some action. It includes the BUTTON operation. The interrupt may be;

1. hardware function key pressed - number of key given
2. specified number of frame holds (i.e. end-of-frame signals) received - this may be used as a delay of specified length, see SKETCH
3. lightpen may "see" an item - number of item given
4. character typed on keyboard of teletype - internal code given. This may effectively be used to expand the number of hardware function keys, and thus the number of actions - see ARCJW024.

ADVANCE (System software)

Causes one character to be read in integer form from the specified peripheral.

ALIGNED (System software)

Used on output to specify the number of places before and after the decimal point to be output for a real number.

ARC (EDGAR)

Generates an arc or circle starting at the current position of the beam

ARCCO

Coordinates of the centre (X,Y,Z), radius R and the polar angles of the beginning and end of the arc may either be read from paper tape or specified in the parameters. Dimensions are in cm, but are converted into plotter increments. A formal parameter list of six elements is set up.
The arc is named for future reference.

ARCSYM

In this case the centre is defined by a symbolic name which has been previously defined, and the radius, start and end polar angles for the arc are defined explicitly. A formal parameter list of six elements is set up, the first three being copied from the formal parameter list for the centre.
The arc is named for future reference.

BUFFER (System software)

Makes available the contents of the peripheral buffer to an *if* clause.

BUTTON (EDGAR)

Waits for attention to be called by the pressing of a single hardware function key, then notes the number of this key. One of the very useful class of instructions which hold up the program to await some action. May be used as a switch to direct subsequent actions, in conjunction with *goto*.

CENCHARACTER (System software)

Plots predefined symbol at the current pen position:

- 1 +
- 2 diamond
- 3 X

plus seven other symbols.

CHAROUT (System software)

Outputs characters in internal code to the specified peripheral.

Must be terminated by CHARCLOSE before the program ends. OUTPUT is the equivalent routine for the display.

CIRCCO

Circle coordinates are specified explicitly. The coordinates for the centre (X,Y,Z), and the radius may be read from paper tape or specified as parameters. Uses ARCCO with start and end polar angles of 0 and 360 degrees.

The circle is named for future reference.

CIRCSYM

In this case the centre of the circle is defined by means of a symbolic name. The remaining parameter R may be read from paper tape or specified as a parameter. Uses ARCSYM, with start and end polar angles of 0 and 360 degrees.

The circle is named for future reference.

CTPTSSR (Circle Through Points and of Specified Radius)

Two points are specified explicitly either by reading from paper tape or by parameters, and a circle of specified radius is drawn through these two points. Of the two possible circles, the required one is selected by specifying the direction of the centre, whether x positive, x negative, y positive or y negative direction, as convenient. The polar angles of the two points are recorded in the formal parameter list.

The circle is named for future reference.

DDRAWF

Draws a figure on the display unit. Expects an array of coordinates with termination codes as created by READFIG.

DECIN

Accepts a decimal number typed on the teletype and terminated by a full stop. The number is converted to binary. This is useful for specifying explicit coordinates on the display screen.

DECINT

Converts a decimal number in character form to binary. The characters are in an array; the starting index and the number of decimal digits must be specified.

DECODE (System software)

Decodes the contents of the peripheral buffer to internal code in integer form. Used with ADVANCE.

DELETE (EDGAR)

Deletes an item from the display file.

DIGITS (System software)

Used on output to specify the number of integer digits to be printed, prefixed by a sign.

DMACIN (System software)

Reads d-Mac coordinates of the form Xnnnn Ynnnn, terminated by halt code. Stores the coordinates in integer arrays.

DRAW (EDGAR)

Copies the code currently in the display buffer into the display file, which causes the item to be displayed. Gives facilities for naming the item and specifying its position (i.e. sequential, or following a predefined item) in the display file.

DRAWA

Draws an arc. Translates and scales the figure as necessary.

Starts off at the first polar angle. Checks that this point and each subsequent point is within the diagram limits. Uses SIN and COS routines to generate the curve, and the arc is drawn in an anticlockwise sense until the second polar angle is reached.

DRAWC

Draws a circle. Translates and scales the figure as necessary.

Starts off at a point on the circle (defined by symbolic name) to the left of the centre, i.e. at point $(x_c - r, y_c)$. Checks that this point and each subsequent point is within the diagram limits.

X coordinates are incremented by 2 plotter increments on each pass and Y coordinates calculated by $y^2 = r^2 - x^2$, where y and x are coordinates relative to the centre of the circle. The circle is drawn in an anticlockwise sense.

DRAWFIG

Draws a standardised pot (created by READFIG) to a specified scale, which may include actual size. Output to the plotter. Only the inner and outer left-hand profiles are carried in the arrays. {The right-hand profile is generated outside this subroutine by reflecting the left-hand outer profile with suppression of any re-entrant footrim. A vertical centre line and horizontal rim line are added. The diagram is labelled by OUTSTRING}. Any figure using the start, end and smoothing codes developed for straight-line figures (e.g. Masons' marks) may be drawn by this subroutine.

DRAWL

Draws a line. Translates and scales the figure as necessary. Checks that the beginning of the line is within the diagram limits. Moves the pen to the beginning of the line. Checks that the end of the line is within the diagram limits, and if so draws the line.

DRAWLINE (System software)

The plotter pen draws a line from the present pen position to the specified location. The pen is left in the pen-down position.

DRAWMAP

Draws a map previously read by READMAP and recorded on magnetic tape. The map identifier, first 3 letters MAP then optional characters terminated by *, is read from paper tape, and the corresponding map outline, if present on the magnetic tape specified, is plotted. The scale is also specified on paper tape, and the map is plotted using a central origin to the required scale. It is the responsibility of the user to ensure that the map will fit on the plotter paper - if it will not, the routine does not draw part of the map, truncating the outline near the edge of the paper at the intersection with the border. A border is in all cases drawn round the complete or truncated map, to A4 size if the map is smaller than A4, or as large as necessary. A scale is drawn below the furthest south point of the map, and a grid north point beyond the furthest east or furthest west point of the map, depending on the room available. Finally the legend is added, consisting of a specified number of strings read by SCRIPT, and therefore allowing upper or lower case. This title is added above the furthest north point of the map, and the border is drawn to allow room for the title.

FILTER

A simple filtering routine used to enhance anomalies; in this case the readings are squared and then divided by 20. Designed especially for the Banjo meter.

FILOT (DES System software)

Copies the display file to the plotter. Runs out the plotter paper.

GETSTRING

Uses INSTRING to input a string of characters punched between string quotes. The packed string, four characters per 24-bit word, is unpacked to internal code in integer form using a (NEAT) machine-code routine. The length of the unpacked code is measured.

GRID

Draws a site grid. The symbolic name of the bottom left-hand corner point of the grid is given, and the angle at which the grid is inclined to the positive x axis. The interval value is given in cm and the number of x and y increments in the grid. Beginnings and ends of all lines of the grid are checked to be within the diagram limits.

HISTOGRAM

Plots a histogram. If parameter DISCRETE =0 then identifying strings and data are read from PT or N items are taken from ARRAY. The items are allocated to ranges determined by parameters FSTVAL, INC and SECVAL, i.e. FSTVAL, FSTVAL:INC:SECVAL and SECVAL. These parameters may have been produced by PREHIST. Counts of the number of items in each interval are produced in array COUNT, and printed. If parameter DISCRETE is non-zero then discrete values are taken direct from array COUNT.

The x-axis of the histogram may be suppressed if desired. If drawn, it is labelled with a description of the meaning of the figures, e.g. WEIGHT IN GRAMS, and the values of the various interval blocks. The height of the histogram blocks is scaled automatically.

Alternatively, if parameter YFIXED is non-zero, the increments of the y axis are specified by parameter YINC. The actual size of one increment on the x and y axes is given in cm by parameters. y scales are labelled on both sides of the histogram, then the histogram blocks are plotted. "Trace" readings, i.e. non-zero counts less than

a specified small value cause a special symbol, + if the x-axis is suppressed, otherwise X, to be plotted instead of a histogram block. The histogram is labelled with the specified strings, and the paper is run out.

INCARD

Reads a card, character by character, storing it in integer character form, and ignoring the 02 (newcard) character.

INCDI

Reads a card image from magnetic tape, or a tape mark.

INSBL

Reads and checks a tape mark and a following SBLO block head.

INSTRING (System software)

Reads a string of characters enclosed in string quotes. The string is stored in packed form, except the opening string quote, four characters per 24-bit word, including case shift and normal characters.

INTNUM

Positions the pen at the defined location, scales the character size according to the current scale factor, then uses WAY to define the character size. Prints the integer decimal number to the specified number of digits at the defined location.

LEADZERO (System software)

Allows any character to replace normally-suppressed leading zeros in an integer number.

LIGHTPEN (EDGAR)

Determines the item number of the item which has been "seen" by the lightpen.

LIMITS

Checks that any specified point is within the defined limits of the diagram. Used for plotter routines in ARCJW013.

LINECO

Specifies line coordinates explicitly. The coordinates are in cm but are converted to plotter increments. A formal parameter list of six elements is set up, specifying the beginning and end points of the line.

The line is named for future reference.

LINESYM

Specifies line coordinates using symbolic names for the beginning and end points. A formal parameter list is set up, being copied from the formal parameter lists of the two specified points.

The line is named for future reference.

MAPGBI

A custom-built procedure which draws a map of GB and Ireland as generated by ARCJW007. Each call of the procedure may draw a new map or use an existing map drawn by a previous call of the procedure.

If a new map is drawn, a title consisting of a variable number of strings is added, together with north point, border and scale.

Points may be plotted on the map at specified NGRs (consisting of two letters plus six figures) with various plotting symbols; +, diamond, X, dot, enhanced dot, doubly-enhanced dot. The Irish grid, which is tilted with respect to the British grid, is used for points within Ireland. The map specification is read from paper tape,

while the map coordinates are read from magnetic tape. Alternatively, the grid references may be taken from an internal array.

Now superceded by DRAWMAP and PLOTREFS.

MAPSTAFFS

A custom-built procedure which draws a map of Staffordshire with a border, grid north point and scale. This program has similar facilities to MAPGBI.

It is now superceded by DRAWMAP and PLOTREFS.

MOVE (EDGAR)

Searches for the first word of the specified item, and if it is a point, replaces it with a new point at the specified coordinates, thus moving the whole of the item on the screen. For this reason it is wise to make sure that the first word of every item is a point command, and it is also good practice to make the second word a set parameter command. These design points have been observed in all EDGAR programs written for this thesis.

MOVEPEN (System software)

The pen is moved in the pen-up position to the specified location. The pen is then left in the pen-up position.

MTCLOSE (System software)

Terminates the magnetic tape correctly.

MTCOND (System software)

Reports on the result of the last magnetic tape operation, e.g. actual length of block read, tape mark etc.

MTDEST (System software)

Opens the magnetic tape mounted on a specified handler as a destination tape, and if it is not protected, checks that its identifier is as specified, and sets its protection status as specified.

MTMARK (System software)

Writes a tape mark .

MTREAD (System software)

Reads the next block from the magnetic tape mounted on the specified

handler and records the information in a specified array from a specified element onwards and not beyond a specified length.

MTREWIND (System software)

Rewinds the magnetic tape.

MTSEEK (System software)

Moves the magnetic tape to the gap following the next tape mark.

MTSOURCE (System software)

Opens the magnetic tape mounted on the specified handler as a source tape, checking that the tape starts with the standard labels, and that its identifier is as specified.

MTWRITE (System software)

Writes a block to the magnetic tape from the specified array, starting at the specified element and continuing for the specified length.

NEWBUF (EDGAR)

Opens a new buffer in the specified array.

NOF

A Boolean procedure which searches the named Boolean array and takes the value TRUE if at least n of the m elements are *true*. This is a useful function in information retrieval, and NOF may be used in the optional Boolean check on magnetic tape records (see ARCJW011).

O(A,B,C,D)

Converts the parameters A,B,C, and D from octal to binary, then packs all four into a single 24-bit word using a (NEAT) machine-code routine. This procedure is used in the specification of new character sets for the plotter (or display). See OCRB.

OCRB

Used to plot in the specially-designed OCRB font instead of the system-defined font. Operates in a similar manner to OUTSTRING. The design of the OCRB characters is based on the NPL character set

(Heap and Laws, 1968). The individual characters are specified as groups of octal numbers, each specifying coordinates of points in the character outline, the moving of the pen in pen-up position, or the drawing of elements above or below the normal grid (4 x 7). Procedure 0 is used to pack these octal numbers. Orientation and size parameters are used to give similar facilities to those supplied by WAY. An italic facility is also available, which applies a progressive displacement from the bottom to the top of a character, and the degree of this slope is determined by the value of the italics parameter.


```

"PROCEDURE"OCRB(STRING,M,ORIENT,SIZE,ITALIC,STARTX,STARTY);
"COMMENT"PLOTS OCRB FONT IN SIMILAR MANNER TO OUTSTRING.
<ORIENT> AND <SIZE> GIVE ORIENTATION AND SIZE PARAMETERS AS FOR
PROCEDURE WAY. AN ITALIC FACILITY IS AVAILABLE IF <ITALIC> IS 1:
(ITALIC MUST OTHERWISE BE 0);
"VALUE"ORIENT,SIZE,ITALIC,STARTX,STARTY;
"INTEGER"M,ORIENT,SIZE,ITALIC,STARTX,STARTY;
"INTEGER""ARRAY"STRING;
"BEGIN"
"INTEGER"I,SPACE,P,Q,SHIFT,PLOTCT,J,K,YINC,DRAW;
"INTEGER""ARRAY"AC[0:639],CHARS,POINTS[1:4];
"SWITCH"PLOT:=PLOT0,PLOT1,PLOT2,PLOT3;
SPACE:=0(77,0,0,0);
"FOR"I:=0"STEP"1"UNTIL"639"DO"AC[I]:=SPACE;
"COMMENT"NUMERALS;
AC[80]:=0(10,2,5,17);AC[81]:=0(37,45,42,30);AC[82]:=0(10,77,0,0);
AC[85]:=0(20,27,16,77);
AC[90]:=0(40,0,1,44);AC[91]:=0(46,37,17,6);
AC[95]:=0(1,10,30,41);AC[96]:=0(43,34,14,47);AC[97]:=0(7,77,0,0);
AC[100]:=0(30,34,70,42);AC[101]:=0(2,27,77,0);
AC[105]:=0(0,30,41,43);AC[106]:=0(34,4,7,47);
AC[110]:=0(37,3,1,10);AC[111]:=0(30,41,43,34);AC[112]:=0(14,3,77,0);
AC[115]:=0(10,12,24,46);AC[116]:=0(47,7,77,0);
AC[120]:=0(14,3,1,10);AC[121]:=0(30,41,43,34);AC[122]:=0(14,5,6,17);
AC[123]:=0(37,46,45,34);
AC[125]:=0(10,44,46,37);AC[126]:=0(17,6,4,13);AC[127]:=0(33,44,77,0);
"COMMENT"UPPER CASE LETTERS;
Similar specifications for the upper and lower case letters, mathematical symbols,
punctuation marks and special signs. A subset may be used if desired.
"COMMENT"END OF FONT SPECIFICATION;
SHIFT:=PLOTCT:=0;
NEXT:P:=STRING[M];M:=M+1;
"FOR"I:=1"STEP"1"UNTIL"4"DO""BEGIN"
"CODE"      % GET      $ P
            % AND:L    $ 63
            % ST       $ Q;
CHARS[I]:=0;"END";
"FOR"I:=1"STEP"1"UNTIL"4"DO""BEGIN"
"IF"CHARS[I]=62"THEN""BEGIN"SHIFT:=1;"GOTO"CH;"END";
"IF"CHARS[I]=63"THEN""BEGIN"SHIFT:=0;"GOTO"CH;"END";
"IF"CHARS[I]=32"AND"SHIFT=1"THEN""GOTO"FIN;
YINC:=DRAW:=0;
"FOR"J:=0"STEP"1"UNTIL"4"DO""BEGIN"
P:=AC((CHARS[I]+64*SHIFT)*5+J);
"FOR"K:=1"STEP"1"UNTIL"4"DO""BEGIN"
"CODE"      % GET      $ P
            % AND:L    $ 63
            % ST       $ Q;
POINTS[K]:=0;"END";
"FOR"K:=1"STEP"1"UNTIL"4"DO""BEGIN"
P:=POINTS[K]"DIV"8;Q:=POINTS[K]-P*8;
"IF"P=5"THEN""BEGIN"YINC:=0;"GOTO"PT;"END";
"IF"P=6"THEN""BEGIN"YINC:=-Q;"GOTO"PT;"END";
"IF"P=7"THEN""BEGIN"DRAW:=0;"IF"Q=7"THEN""GOTO"CH1"ELSE""GOTO"PT;"END";
P:=(P+0.2*ITALIC*Q+PLOTCT*5)*SIZE;Q:=(Q+YINC)*SIZE;
"GOTO"PLOT[ORIENT+1];
PLOT0:"IF"DRAW=0"THEN"MOVEPEN(STARTX+P,STARTY+Q)
"ELSE"DRAWLINE(STARTX+P,STARTY+Q);"GOTO"PT1;
PLOT1:"IF"DRAW=0"THEN"MOVEPEN(STARTX-Q,STARTY+P)
"ELSE"DRAWLINE(STARTX-Q,STARTY+P);"GOTO"PT1;
PLOT2:"IF"DRAW=0"THEN"MOVEPEN(STARTX-P,STARTY-Q)
"ELSE"DRAWLINE(STARTX-P,STARTY-Q);"GOTO"PT1;
PLOT3:"IF"DRAW=0"THEN"MOVEPEN(STARTX+Q,STARTY-P)
"ELSE"DRAWLINE(STARTX+Q,STARTY-P);
PT1:DRAW:=1;
PT:"END";"END";
CH1:PLOTCT:=PLOTCT+1;
CH:"END";
"GOTO"NEXT;
FIN:"END"OCRB;

```

Figure A1

Coding for OCRB font specification

OLDSET (EDGAR)

Allows an old buffer to be reopened, without deletion of code already assembled in the buffer.

ORIGIN

Positions the pen at the new origin, then uses SETORIGIN to define the new plotter origin.

OUTCBI

Writes a card image to magnetic tape with an optional

OUTPUT (EDGAR)

Generates a character on the screen of the plot and printer specified by SETP. Can be used with OUTCBI as a subcommand.

OUTSBL

Outputs a tape mark and special black band (used for tape marks).

OUTSBL (System software)

Outputs a string of characters plotted with a pen or printer.

stored in a standard format, four characters per line.

plotter or printer, and the output is stored in a standard format.

and the output is stored in a standard format.

tive

PERM

Deter

cross

PERM

A recursive procedure to process the output of the plotter.

outputs the intention is to show the output of the plotter.

according to the state of the plotter.

a defined integer

Figure A2

Sample outputs of the OCRB font, showing upper case letters, numerals, mathematical symbols, punctuation marks and special signs, both in normal and italic modes. The set shown corresponds to the ICL 4130 internal character set. Use of the orientation feature and varying degrees of italics are shown in the upper part of the figure.

OLDBUF (EDGAR)

Allows an old buffer to be reopened, without deletion of code already assembled in the buffer.

ORIGIN

Positions the pen at the new origin, then uses SETORIGIN to define the new plotter origin.

OUTCDI

Writes a card image to magnetic tape with an optional tape mark .

OUTPUT (EDGAR)

Generates a character on the screen of the size and brightness specified by SETP. Can be used with DECODE as a parameter.

OUTSBL

Outputs a tape mark and special block head SBLO to magnetic tape.

OUTSTRING (System software)

Outputs a string of characters previously input by INSTRING (i.e. stored in packed format, four characters per 24-bit word). On the plotter both case normal and case shift characters are operable, and cause the output of upper case and lower case characters respectively, a fact that is not made clear in the software manuals.

PENTRAK (EDGAR)

Determines the coordinates currently indicated by the tracking cross, i.e. the last known position of the light pen.

PERM

A recursive procedure to permute output procedures. For graphical outputs the intention is to draw these in different locations according to the state of the permutation. Indices are taken from a defined integer array, the number of such indices being defined in parameter ORDER. In the initial call DEPTH = ORDER, but DEPTH is used to control the recursive calls of PERM. Data cards give the

required output routines, using ALGOL statements and procedure calls. The depth of recursion is governed by the value of ORDER. Used for Dr F.Celoria (Strike-a-lights).


```

"PROCEDURE"PERM(INDEX,DEPTH,ORDER);
"COMMENT" A RECURSIVE PROCEDURE TO PERMUTE OUTPUT PROCEDURES.
FOR GRAPHICAL OUTPUTS THE INTENTION IS TO DRAW THESE IN DIFFERENT
LOCATIONS ACCORDING TO THE STATE OF THE PERMUTATION. INDICES ARE
TAKEN FROM INTEGER ARRAY <INDEX> FOR THE PERMUTATION, THE NUMBER OF
SUCH INDICES BEING DETERMINED BY THE VALUE OF <ORDER>. IN THE
INITIAL CALL <DEPTH> = <ORDER>, BUT <DEPTH> IS USED TO CONTROL THE
RECURSIVE CALLS OF PERM;
"VALUE"ORDER,DEPTH;
"INTEGER"ORDER,DEPTH;"INTEGER""ARRAY"INDEX;
"BEGIN"
"INTEGER"J;
"SWITCH"FIG:=FIG1,FIG2,FIG3,FIG4;
"FOR"INDEX[DEPTH]:=1"STEP"1"UNTIL"ORDER"DO""BEGIN"
"IF"DEPTH=ORDER"THEN""GOTO"NEXT;
"FOR"J:=ORDER"STEP"-1"UNTIL"DEPTH+1"DO"
"IF"INDEX[DEPTH]=INDEX[J]"THEN""GOTO"REJ;
NEXT:"IF"DEPTH=1"THEN""BEGIN"
"COMMENT"OUTPUT STATEMENTS;
"FOR"J:=ORDER"STEP"-1"UNTIL"1"DO"
"BEGIN"
REALX:=0.0;REALY:=-6.0+4.0*(J-1);
"GOTO"FIG[INDEX[J]];
FIG1:DRAWFIG(X1,Y1,I1,REALX,REALY,0.2);"GOTO"COMP;
FIG2:DRAWFIG(X2,Y2,I2,REALX,REALY,0.2);"GOTO"COMP;
FIG3:DRAWFIG(X3,Y3,I3,REALX,REALY,0.2);"GOTO"COMP;
FIG4:DRAWFIG(X4,Y4,I4,REALX,REALY,0.2);"GOTO"COMP;
COMP:"END";
ORIGIN(4.0,0.0);
"COMMENT"END OF OUTPUT STATEMENTS;
"END""ELSE"
PERM(INDEX,DEPTH-1,ORDER);"COMMENT"RECURSIVE CALL;
REJ:"END";
"END"PERM;

```

Figure A3

Coding for PERM

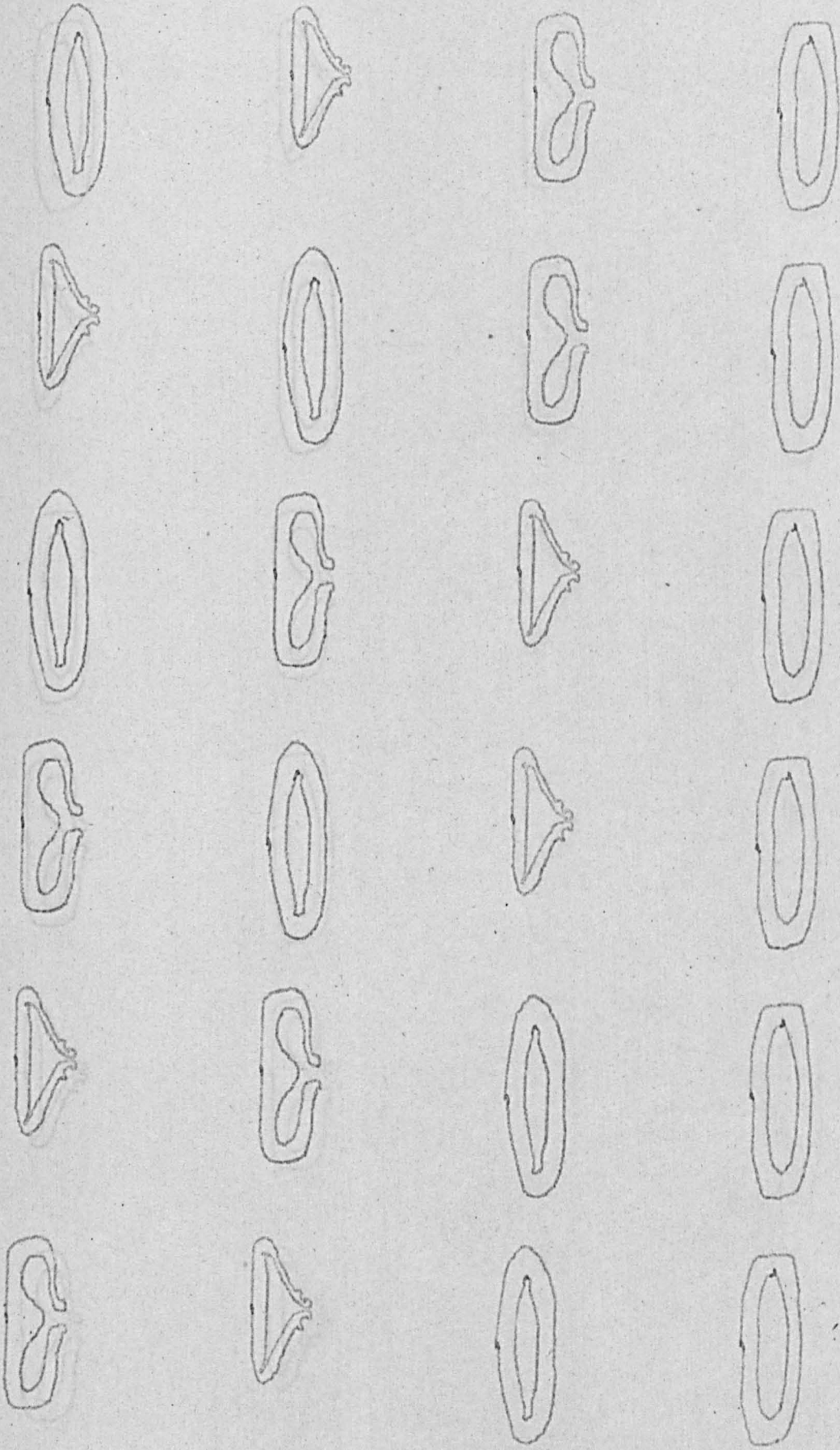


Figure A4

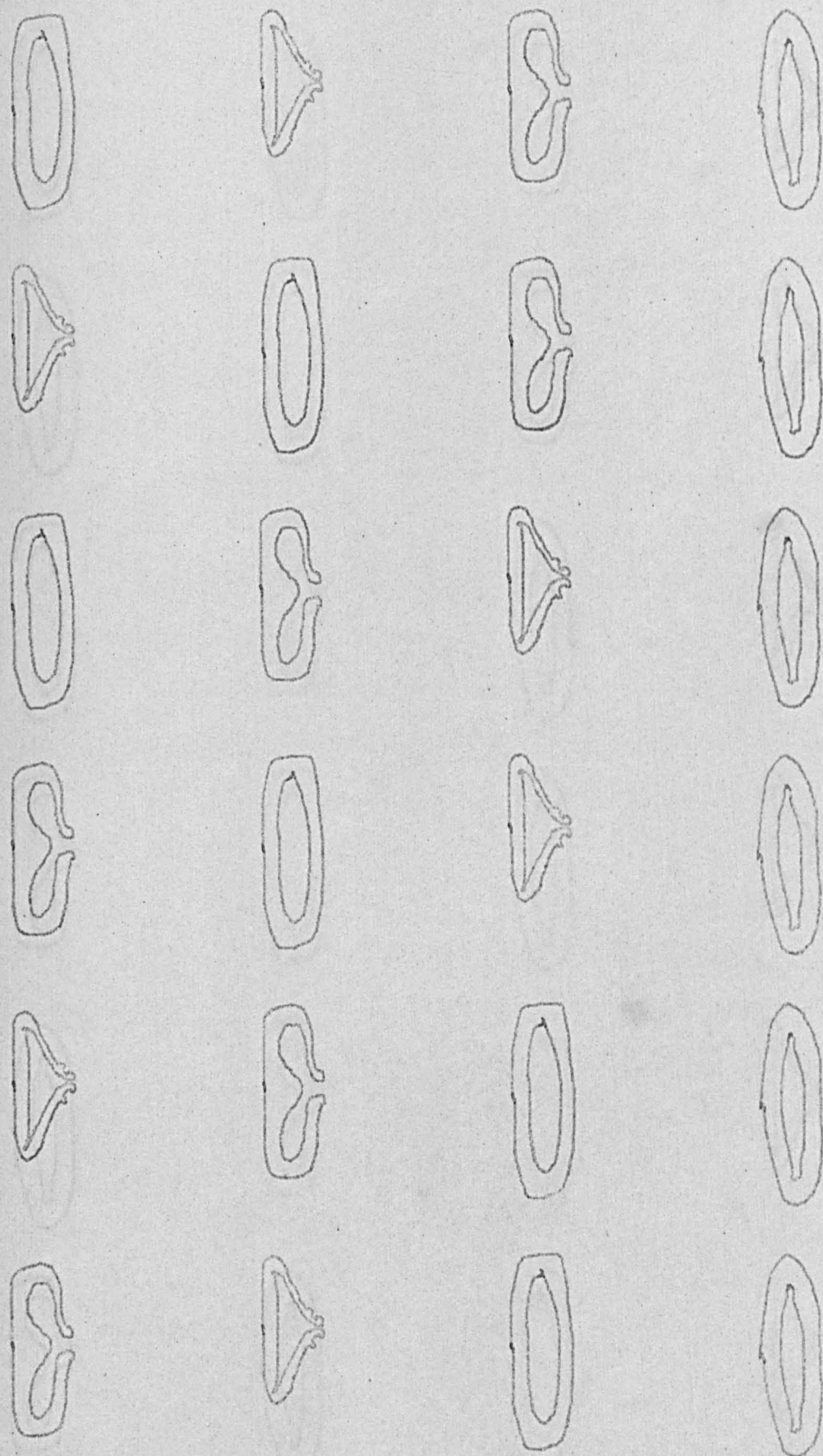


Figure A4

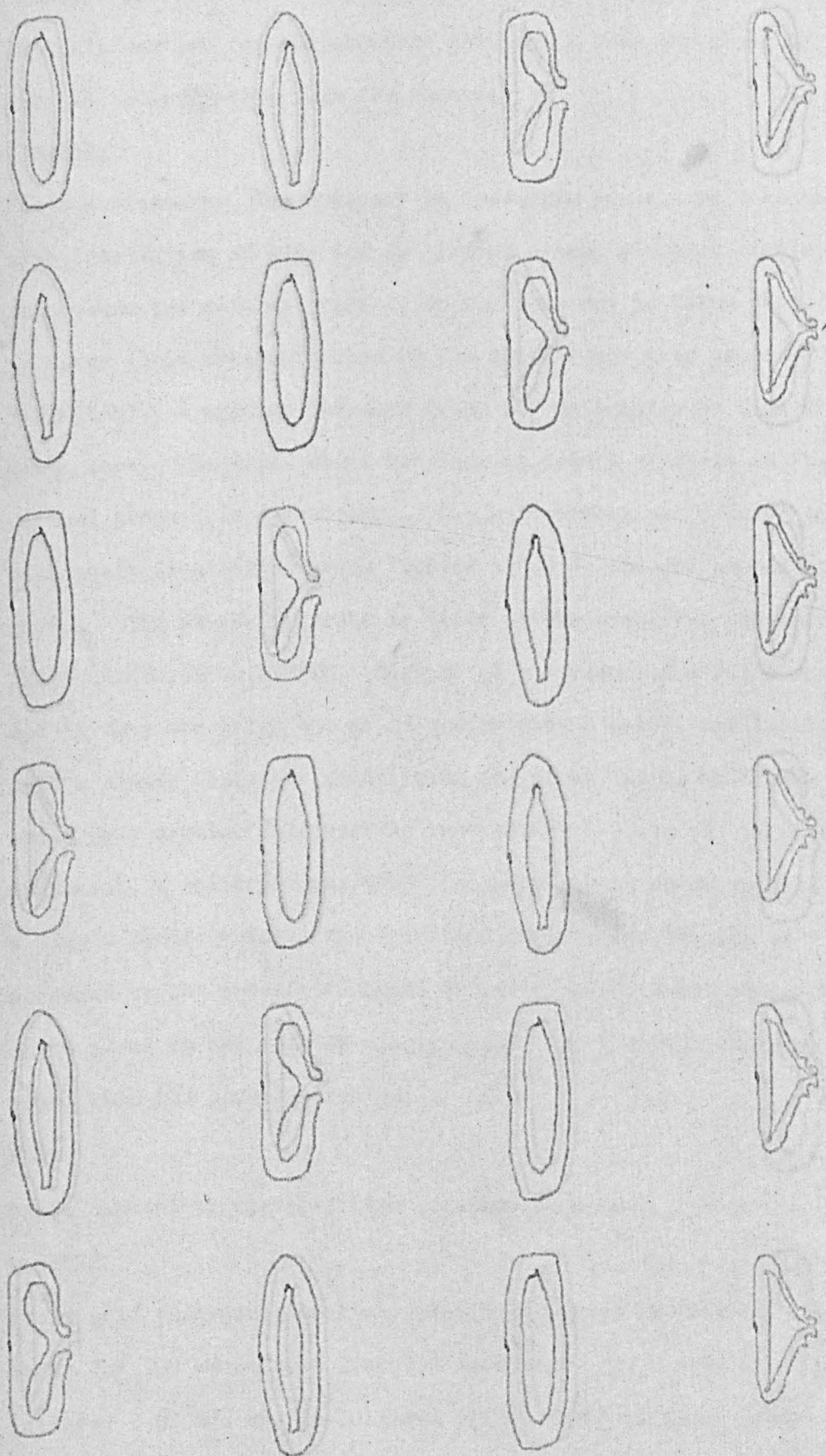


Figure A4

PIECHART

Performs perspective calculations for the display, assuming an initial location for the observer (30 miles from the plane of the horizon) and adjusting this for zooming.

PIECHART

as a piechart. The data may be read from paper tape together

with identifying strings and the radius of the piechart circle, in

case the data is printed; or the data may be taken direct from

array (this option is used if the radius has been provided by

PIECHART). A caption for each group may optionally be read from

paper tape. The data, which consists of counts of items in the

various groups, is percentage. The percentages are printed together

with their identifying group letters (A-Z), and the number of

groups. The piechart circle is drawn at the specified radius using

PIECHART, CIRC, ARCCO, LIMITS. Sectors of the required angle are drawn,

and if they are large enough to accommodate a label, are labelled

with a single character identifying the group (using OUTLINE pre-

pared by a machine code packing routine which packs the required

character, a shift character, closing string quote and space into

a single 24-bit word. The specified strings are printed as a title,

preceded by the phrase 'PIECHART OF'. If the optional captions are

added to the plotted diagram, they are plotted below the dia-

gram then the paper is run out.

Plots a point at the specified location as a small square.

PIECHART

Plots grid references on a map previously drawn by PIECHART. Grid

references for 100 km squares are read from paper tape: each is two

letters, e.g. SJ, and is followed by a space. Specified by

PIECHART. References, in km east and north from the origin, are

PIECHART

Figure A4

PERS

Performs perspective calculations for the display, assuming an initial location for the observer (30 inches from the plane of the screen) and adjusting this for zooming.

PIECHART

Plots a piechart. The data may be read from paper tape together with identifying strings and the radius of the piechart circle, in which case the data is printed; or the data may be taken direct from an array (this option is used if the routine has been preceded by HISTOGRAM). A caption for each group may optionally be read from paper tape. The data, which consists of counts of items in the various groups, is percentaged. The percentages are printed together with their identifying group letters A-H, J-Z, and the number of groups. The piechart circle is drawn to the specified radius using DRAWC, CIRCCO, ARCCO, LIMITS. Sectors of the required angle are drawn, and if they are large enough to accommodate a label, are labelled with a single character identifying the group (using OUTSTRING preceded by a machine code packing routine which packs the required character, a shift out character, closing string quote and space into a single 24-bit word). The specified strings are printed as a title, preceded by the phrase 'PIECHART OF'. If the optional captions are to be added to the plotted diagram, they are plotted below the diagram, then the paper is run out.

PLOT

Plots a point at the specified location as a small square.

PLOTREFS

Plots grid references on a map previously drawn by DRAWMAP. Valid codes for 100 km squares are read from paper tape: each is of two letters, e.g. SJ, and is followed by two integers specifying the distances, in km east and north from the origin of the SW corner of

the square. Alternatively, a single letter followed by *, e.g. N*, and two integers indicates that a whole 500 km square is defined, the second letter for each constituent 100 km square following the convention for the British National Grid. ** followed by a non-zero real number and two integers indicates that a further grid is defined aligned at an angle to the first grid specified in radians by the real number, its origin displaced distances in km east and north from the first origin. This feature allows e.g. the Irish grid to be accommodated with the British National Grid. The definition of squares within the new grid then continues, using distances in km parallel to the base and parallel to the upright of the new grid. ** followed by the number 0 causes termination of the definitions. Next four integers are read from paper tape specifying the east range (maximum then minimum) and the north range (maximum then minimum) in km of valid Grid references from the first origin. Finally, grid references, each of two letters and six figures are read from paper tape or from core, checked for validity, and plotted with the symbol specified by * and a single digit: 1 +, 2 diamond, 3 X, 4 dot, 5 enhanced dot, 6 doubly-enhanced dot, *0 causes runout of the plotter paper and termination, while ** in the symbol specification mode causes the shade mode to be entered (see SHADE). In the shade mode the type of shading is defined by * and a single digit: 1-6 as above, 7 vertical line, 8 positive diagonal, 9 horizontal line, A negative diagonal ("A" is hexadecimal for 10). The area to be shaded is outlined by a string of grid references. In the shade mode ** causes the plotting mode to be entered, while *0 causes runout of the plotter paper and termination of the routine.

POINT (EDGAR)

Specifies a point at the given coordinates. The point may either be invisible or visible.

POINT (ARCJW013)

Specifies point coordinates explicitly. The coordinates may be read from paper tape or specified as parameters. Dimensions are in cm but are converted to plotter increments. A formal parameter list is set up for the point, which is named for future reference.

POSITION

Accepts coordinates in cm. Scales them according to the current scale factor, then checks if they are within the diagram limits (uses LIMITS) - if not makes an exit to the named label. Otherwise positions the pen in pen-up position at the desired location.

PREHIST

Adapts the raw data so that HISTOGRAM can be used as a 'lazy man's histogram', i.e. without any specification of ranges, amount of data, or interval by the user. The routine expects a terminating value

1000000. The items are counted, stored and printed. The maximum, minimum and range are calculated. The interval is chosen as a function of the range, and the range is adjusted to cover an integral number of intervals. These parameters are recorded for use in HISTOGRAM.

"PRINT" (System software)

Output command, applicable to all output peripherals.

PTNUMS

Displays the numerical identifiers of all points in the point list at bright illumination. Exits when a point identifier is indicated by a lightpen hit or by typing the numerical identifier on the teletype.

PUNCH (System software)

Used to specify the peripheral for output:

- 1 Paper tape punch
- 3 Teletype
- 4 Line printer
- 5 Digital incremental plotter
- 9 Graphical display

RANNO

A pseudo-random number generator, based on the remainder of an integer division. Modulus = remainder, sign -ve if the 1s bit is 1, +ve if the 1s bit is 0. Before exit the next pseudo-random number to be submitted to integer division on the next pass is generated by adding the present number to a copy of itself shifted right 3 binary places.

"READ" (System software)

Input command, applicable to all input peripherals.

READER (System software)

Specifies the input peripheral:

- 1 Paper tape reader
- 3 Teletype
- 6 Card reader

READFIG

Reads the name of a pot (using INSTRING), the scale factor, then d-Mac coordinates of the inner and outer profiles; rotates, smoothes and scales the coordinates to standard height, with adjustment of the scale factor. The result is a figure aligned to the true vertical and of standardised height, in a form that may be directly written to magnetic tape. The following codes have been developed to describe straight-line figures, or curves built up of short straight lines; 0000 starting end; 0001 finishing end; 9000 starting intersection; 9001 finishing intersection; 9999 intersection which is part of a continuous line; 0002 angle in a closed figure. The following codes are used to control smoothing, which is undesirable in the case of sharply-angled figures, e.g. rectangles; 0003 switch off smoothing; 0004 switch on smoothing. These codes have been used successfully in the description

of figures, e.g. Masons' marks, clay pipes, pottery, especially the first, which tend to be simple straight-line figures.

READMAP

Reads from paper tape the following information:

the character identifier of a map terminated by *;

the scale of the original map;

the scale of the desired reproduction on magnetic tape;

d-Mac coordinates for the grid origin and a point to the grid north of the origin, followed by the halt code;

a complete set of data for each detached line or closed curve on the map as follows:

for other than the first line, a non-zero integer (a zero integer in this place indicates that the map is complete;

offsets in km east and north for the sub-origin of this line

relative to the origin (this is to allow map insets to be traced);

d-Mac coordinates for the sub-origin and up to 3000 desired points on the line, followed by the halt code;

N.B. if offsets are zero, the origin is taken as the sub-origin

and the origin coordinates must be specified in place of the sub-origin coordinates.

The procedure rotates the map to a truly vertical alignment, x-axis to grid north, y-axis to grid west, translates all coordinates to a zero origin and scales them as required. Finally the coordinates are written to the magnetic tape new master, which is copied at the start of the procedure from the old master. The data is preceded on magnetic tape by the map identifier and scale. Another map is then processed (or the procedure terminates if * is input in place of a map identifier). Uses OUTSBL, SEARCH, and DMACIN.

REALNUM

Positions the pen at the defined location. Scales the character size according to the current scale factor, then uses WAY to define

the character size. Prints the real number to the specified number of integral and fractional places at the defined location.

RESET (EDGAR)

Resets the display file and starts the display running. Defines an integer array for the display file and the maximum number of items to be displayed.

ROT

A rotation procedure used to rotate all axes as required on the display. All coordinates are recalculated.

ROTATE

Rotates the axes of the diagram. Alters the rotation parameters so that all action routines for drawing lines, circles, etc. are modified in their absolute positions on the plotter.

SAMELINE (System software)

Used on output to suppress the newline which is otherwise output before each number.

SCALE

Alters the scale for succeeding commands.

SCRIPT

Produces a packed string including lower case alphabets. Uses GETSTRING to input a string of characters punched between string quotes, and to unpack them to integer form. The routine gives the facility to change between upper and lower cases and vice versa at will, using code 1 for shift out to produce lower case and 2 for shift in to produce upper case. A marker indicates the current shift state. Code 1 is replaced by the shift out code and the marker is set. Code 2 is replaced by the shift in code and the marker is reset. Existing shift out and shift in codes remain unchanged and the marker is set to the correct state. If the space code

occurs while the marker is set, the shift in, space, and shift out codes are produced, since space does not exist in shift out. Other codes are unchanged. After the closing string quote is detected, remaining space characters padding the original code to a multiple of 4 are reproduced unchanged. Finally the augmented code is padded to a multiple of 4 with new space characters, and the whole is packed by a (NEAT) machine-code routine; if the last resulting 24-bit word consists of all space characters (because of the augmentation and existing padding) this is taken into account in the reported length. This routine takes advantage of the existing lower case graphics in the software for the plotter, which are otherwise not attainable by OUTSTRING, and are not mentioned in the systems manuals.

SEARCH

Searches for the occurrence of a given sequence of characters in any part of a defined string.

SETORIGIN (System software)

Moves the plotter pen in the pen-up position to a point situated a specified number of steps from the left-hand margin, makes this point the origin for subsequent movements.

SETP (EDGAR)

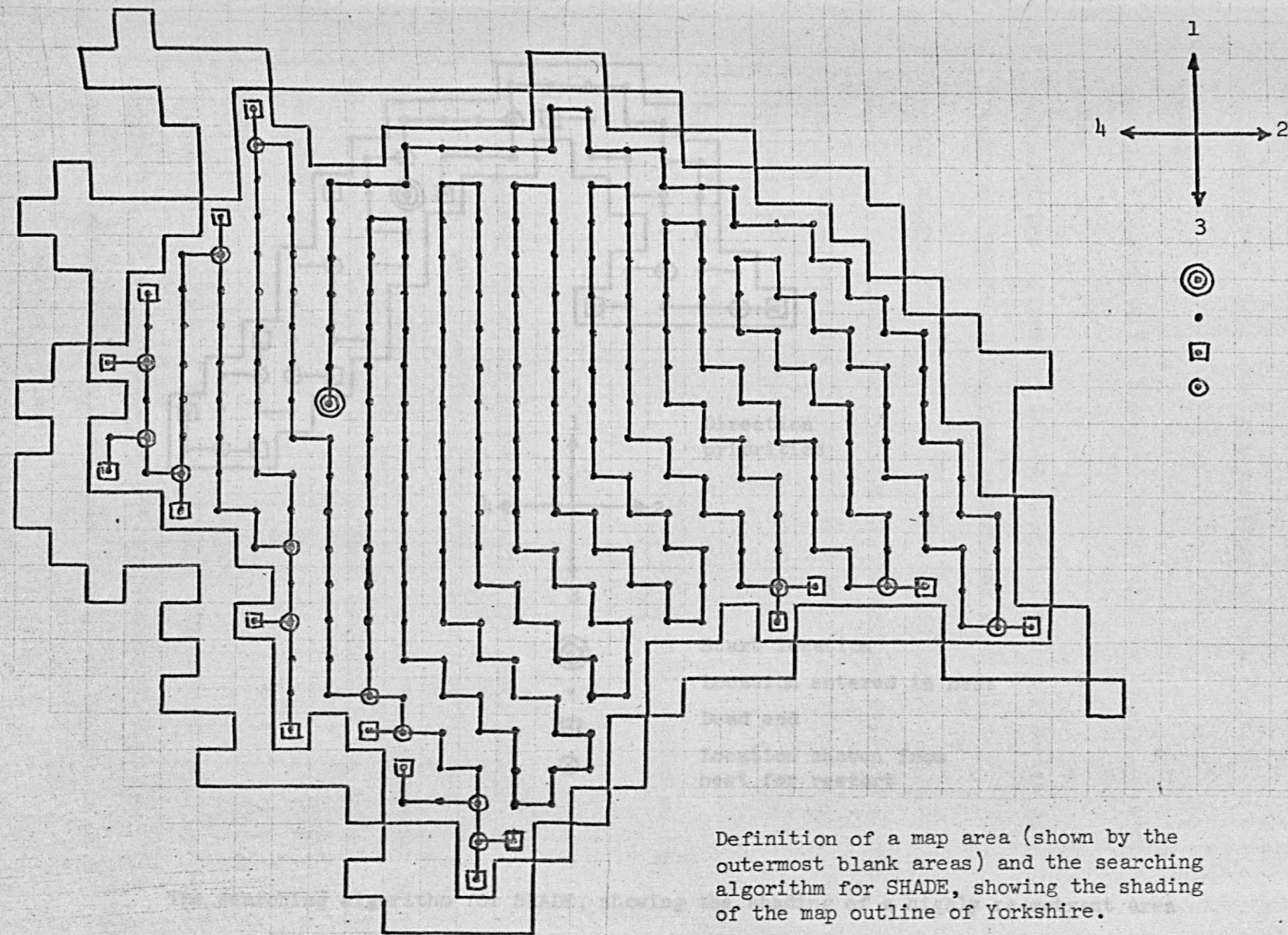
Sets parameters, viz. brightness, character size, edge status and slave screen control. This command is conveniently placed second in all items output to the display file, the first command being a POINT command (to allow the use of MOVE).

SETTRAK (EDGAR)

Either removes or generates the tracking cross, for use with the lightpen in the specification of points on the screen. The tracking cross is part of the code in the display file, and must consequently be deleted before a diagram is plotted in permanent form.

SHADE

Allows closed areas of a map, specified by chains of grid references, to be shaded in a variety of ways. The grid references may be read from paper tape or from an internal array. The following shade elements which build up to form the shading pattern are available: +, diamond, X, dot, and cross-hatching in the vertical, positive diagonal, horizontal and negative diagonal senses. An algorithm similar to a maze-searching routine is defined to determine which of the available Boolean matrix points at 0.1 inch spacing shall be shaded. The matrix points are initially set *false*, then the boundary of the area is defined by a connected series of *true* points. Starting at an interior point, also specified as a map reference, the setting of the Boolean matrix points is controlled by a nesting store until all internal points are set *true*.



Direction
priorities

Start location

Location entered in nest

Dead end

Location chosen from
nest for restart

Definition of a map area (shown by the
outermost blank areas) and the searching
algorithm for SHADE, showing the shading
of the map outline of Yorkshire.

Figure A5

Sheet 1 of 2 sheets

